UNCLASSIFIED

DTIC FILE COPY

AGE

MENTATION PAGE

Form Approved
OMB No. 0704-0188

AD-A223 873

1b. *RESTRICTIVE MARKINGS*	
2a. *SECURITY CLASSIFICATION AUTHORITY*	3. *DISTRIBUTION/AVAILABILITY OF REPORT* Approved for public release; distribution is unlimited.
2b. *DECLASSIFICATION/DOWNGRADING SCHEDULE*	
4. *PERFORMING ORGANIZATION REPORT NUMBER(S)*	5. *MONITORING ORGANIZATION REPORT NUMBER(S)*
6a. *NAME OF PERFORMING ORGANIZATION* Student, HQDA, MILPERCEN	6b. *OFFICE SYMBOL (if applicable)* DAPC-OPA-E
7a. *NAME OF MONITORING ORGANIZATION* HQDA, MILPERCEN,	
6c. *ADDRESS (City, State and ZIP Code)* 200 Stovall Street Alexandria, VA 22332	7b. *ADDRESS (City, State and ZIP Code)* ATTN: DAPC-OPA-E 200 Stovall Street Alexandria, VA 22332
8a. *NAME OF FUNDING/SPONSORING ORGANIZATION*	8b. *OFFICE SYMBOL (if applicable)*
9. *PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER*	
8c. *ADDRESS (City, State and ZIP Code)*	10. *SOURCE OF FUNDING NUMBERS*

PROGRAM ELEMENT NO.	*PROJECT NO.*	*TASK NO.*	WORK UNIT ACCESSION NO.

11. *TITLE (include Security Classification)*
A Design Tool Utilizing Stoichiometric Structure for the Analysis of Biochemical Reaction Networks

12. *PERSONAL AUTHOR(S)*
Captain Stephen E. Kelly

13a. TYPE OF REPORT	13b. TIME COVERED *From*	*To*	14. DATE OF REPORT *(Year, Month, Day)*	15. PAGE COUNT
Final Report			20 May 1990	172

16. SUPPLEMENTARY NOTATION
A Master's Thesis → Computer Programming and Software 4 FJG

17. COSATI CODES			18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)* Theses, Quantitative analysis, Chemical Analysis, Biochemical Reaction Networks, metabolic pathway, Modeling, Programming, Toolbox Program, MINOS, Optimization Package, HyperCard, Apple Computer
FIELD	GROUP	SUB-GROUP	

DTIC ELECTE JUL 13 1990 S GE

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

A method is proposed for the analysis of possible distributions of products in biochemical reaction networks using linear optimization techniques. Software to assist in the conduct of this analysis was developed. Several biochemical systems were analyzed to examine the effects of different constraining situations.

(continued on reverse)

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
[X] UNCLASSIFIED/UNLIMITED [ ] SAME AS RPT. [ ] DTIC USERS	UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE *(Include Area Code)*	22c. OFFICE SYMBOL
Captain Stephen E. Kelly	DSN 789-5637	STEDP-MT-TM-CB

DD Form 1473, JUN 86 *Previous editions are obsolete.* SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

19. ABSTRACT (continued)

Software was developed to simplify and reduce the manual data manipulation needed to conduct this analysis. The software consists of two major components: a specially constructed database and a commercial linear optimization package.

The database program was constructed to store and manipulate information on chemical species and on the reactions involving them. The database contains the elemental composition of chemical species, quantitative information on the reactions, graphic depiction of reactions and species, and general data such as notes and references.

The database can assemble information about many reactions to form a stoichiometric model of a reaction network. The database assembles user-supplied information on system constraints and information from the database to build a linear programming problem file which is then passed to an optimization routine for solution. The solution is recovered and presented in a readily comprehensible format.

The solution provides information on the overall reaction the network performs as well as the relative rates of the individual reactions within the network. The results provide some information about possible yields of products and about the functional nature of biochemical pathways. It also provides a means of checking consistency and closure in fermentation material balances.

Several biochemical reaction networks were examined. A proposed pathway for the biosynthetic production of astaxanthin, a natural red pigment was analyzed. Penicillin production was examined for the effects of various carbon and reduction/oxidation limits. The production of 1,3-propanediol (1,3-PD) was examined with the pathway for 1,3-PD overlaid onto the general metabolic pathways of Escherichia coli.

A DESIGN TOOL UTILIZING STOICHIOMETRIC STRUCTURE FOR THE ANALYSIS OF BIOCHEMICAL REACTION NETWORKS

Captain Stephen E. Kelly
U.S. Total Army Personnel Command (TAPC-OPA-E)
200 Stovall Street
Alexandria, Virginia 22332

Final Report
20 May 1990

Approved for public release; distribution is unlimited

A thesis submitted to the University of Wisconsin - Madison
in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE (Chemical Engineering)

A DESIGN TOOL UTILIZING STOICHIOMETRIC STRUCTURE FOR THE ANALYSIS OF BIOCHEMICAL REACTION NETWORKS

by

STEPHEN EDWARD KELLY

A thesis submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE
(Chemical Engineering)

at the
UNIVERSITY OF WISCONSIN – MADISON
1990

ACCEPTED BY _____

TITLE _____

Abstract

A method is proposed for the analysis of the possible distributions of products in biochemical reaction networks using linear optimization techniques. Computer software tools to assist in the conduct of this analysis were developed. Several biochemical systems, both actual and notional, were analyzed using the software tools to examine the effects of different constraining situations.

Software was developed to simplify and reduce the amount of manual data manipulation needed to conduct this analysis. The software consists of two major components: a specially constructed database program and a commercially available linear optimization package.

The database program was constructed to store and manipulate information on chemical species and on the reactions involving them, both enzymatically catalyzed and non-enzymatic. The database provides for the inclusion of the elemental composition of chemical species, quantitative information on the reactions, graphic depiction of reactions and species, and general data such as notes and references.

The database can assemble information about many reactions to form an accurate stoichiometric model of a reaction network. The database assembles user supplied information on the system constraints and information from the database to build a linear programming problem file which is then passed to a standard commercial optimization routine for solution. The solution to the problem is recovered and presented in a readily comprehensible format.

The solution provides information on the overall reaction the network performs as well as the relative rates of the individual reactions within the network. The linear programming results provide some information about possible yields of products and about the functional nature of the biochemical pathways. It also provides a means of checking for consistency and closure in fermentation material balances.

Several biochemical reaction networks were examined. A proposed pathway for the biosynthetic production of astaxanthin, a natural red pigment was analyzed. The overall reaction system of penicillin production was examined for the effects of various carbon and reduction/oxidation limits. The production of 1,3-propanediol (1,3-PD)

was examined with the pathway for 1,3-PD overlaid onto the general metabolic pathways of *Escherichia coli*.

Acknowledgements

Primary support for this project came from the United States Army which, under the auspices of the the Advanced Civil Schooling Program, provided me the opportunity and financial resources to attend graduate school. Computer hardware was provided by Apple Computer Incorporated through a Project ROTA grant.

I would like to express my appreciation to my faculty advisor, Professor Douglas Cameron, for his guidance and support. The freedom of action and personal initiative he allowed in my work made this project a worthwhile learning experience.

And, of course, I must thank my family for their unflagging support for and tolerance of my efforts. My wife, Angie, and children, Jocelyn and Nicholas, have made this academic sojourn truly enjoyable and productive.

Table of Contents

Abstract iii

Acknowledgements v

Table of Contents vi

List of Figures xi

List of Tables xii

Chapter 1. Summary 1

1.1. The Problem 1

1.2. An Approach to a Solution 2

1.3. Some Case Studies 5

1.4. Conclusions 7

Chapter 2. Literature Review 8

2.1. Biochemical Modeling 8

2.2. Detailed Kinetic Models 9

2.3. Early Simple Models 10

2.4. Pathway Determination 11

2.5. Yield Analysis 13

2.6. Stoichiometry in Fermentation Control 16

2.7. Stoichiometry and the Fermentation Equation 17

2.8. Stoichiometric Data Analysis 19

2.9. Other Uses of Stoichiometry in a Biochemical Network ... 21

Chapter 3. Linear Modeling of Reaction Networks 23

3.1. Quantitative Analysis of Biochemical Systems 23

3.2. The General Approach to Stoichiometric Notation 23
3.3. Biochemical Systems as Open Systems 25
3.4. Linear Programming 27
3.5. Manipulating Inequalities 28
3.6. Constraints for Reaction Types 30
3.7. Constraints on Chemical Species 31
3.8. Objective Functions 35
3.9. Solving the Problem 38
3.10. The Pathway Toolbox Program 39
Chapter 4. Case Study I: Astaxanthin Biosynthesis 53
4.1. Carotenoid Uses 53
4.2. Astaxanthin Production 54
4.3. Biosynthetic Pathway for Astaxanthin 55
4.3.1. Isoprenoid Formation 56
4.3.2. Carotenoid Formation 58
4.3.3. Astaxanthin Formation 60
4.3.4. Other Reactions 63
4.4. Analysis 66
4.4.1. Carbon Limitation 66
4.4.1.1. Constraints on Reactions 66
4.4.1.2. Constraints on Chemical Species 69
4.4.1.3. Results 72
4.4.2. Reduction/Oxidation Limitations 76
4.4.2.1. Constraints on Reactions 76

4.4.2.2. Constraints on Chemical Species 76
4.4.2.3. Results 76
4.4.3. Other Constraints 80
4.5. Summary 82
Chapter 5. Case Study II: Penicillin Biosynthesis 84
5.1. Overview of the Process 84
5.2. Pathway for Penicillin Biosynthesis 84
5.2.1. Cysteine Biosynthesis 85
5.2.2. Valine Biosynthesis 89
5.2.3. α-aminoadipate Synthesis 91
5.2.4. Penicillin Assembly 93
5.2.5. Other Reactions 95
5.3. Yield Analysis 98
5.3.1. Carbon Limitations 98
5.3.1.1. Constraints on Reactions 98
5.3.1.2. Constraints on Chemical Species 100
5.3.1.3. Results 104
5.3.2. Reduction and Energy Limitations 108
5.3.2.1. Constraints on Reactions 108
5.3.2.2. Constraints on Chemical Species 109
5.3.2.3. Results 110
5.4. Summary 114
Chapter 6. Case Study III: Aspects of 1,3-Propanediol Biosynthesis 115

6.1. Overview of the Process 115
6.2. Pathways for 1,3-Propanediol Biosynthesis 117
6.2.1. Pyruvate Formation 117
6.2.2. Reduction to 1,3-Propanediol 120
6.2.3. Pyruvate Dehydrogenase and Partial Tricarboxylic Acid Cycle 122
6.2.4. Reduced Metabolites 125
6.2.5. Pentose Phosphate Pathway 127
6.2.6. Other Reactions 130
6.3. An Analysis of Potential Yields of 1,3-Propanediol 131
6.3.1. Carbon Limited Biosynthesis 131
6.3.1.1. Constraints on reactions 131
6.3.1.2. Constraints on Chemical Species 134
6.3.1.3. Results 138
6.3.2. Reducing Power Limits to Conversion Yield 139
6.3.2.1. Constraints on Reactions 139
6.3.2.2. Constraints on Chemical Species 139
6.3.2.3. Results 140
6.3.3. Pathway Limited Yields 142
6.3.3.1. Constraints on Reactions 143
6.3.3.2. Constraints on Chemical Species 143
6.3.3.3. Results 144
6.4. Summary 146
Chapter 7. Further Work 148

7.1. Objective Functions 148
7.2. User Interface 149
7.2.1. User Interface - Representation 149
7.2.2. User Interface - Setting Constraints 150
Abbreviations 152
References 154

List of Figures

Figure 3.1. Software Structure 40
Figure 3.2. Data Card for Information on a Chemical Species from Pathway Toolbox Program 43
Figure 3.3. Data Card for a Reaction 46
Figure 3.4. Species Constraints 48
Figure 3.5. Reaction Constraints 50
Figure 3.6. Results Example 52
Figure 4.1. Biosynthesis of Isopentenyl Pyrophosphate 57
Figure 4.2. Synthesis of Geranylgeranyl Pyrophosphate 59
Figure 4.3. Carotenoid Synthesis Reactions 61
Figure 5.1. Cysteine Pathway 86
Figure 5.2. Sulfate Reduction 88
Figure 5.3. Valine Pathway 90
Figure 5.4. α-aminoadipate Synthesis 92
Figure 5.4. α-aminoadipate Synthesis 94
Figure 6.1. Pyruvate Formation 118
Figure 6.2. 1,3-Propanediol Synthesis Pathway 121
Figure 6.3. Partial Tricarboxylic Acid Cycle 123
Figure 6.4. Reduced Metabolic Products 126
Figure 6.5. Pentose Phosphate Pathway 128

List of Tables

Table 3.1. Constraints on Chemical Species Coefficients, R_{Si} 33

Table 4.1. Constraints on Astaxanthin Pathway Reactions 68

Table 4.2. Constraints on Astaxanthin Pathway Chemical Species .. 70

Table 4.3. Optimized Reaction Rates for Astaxanthin Production 73

Table 5.1. Reversibilities of Reactions 98

Table 5.2. Constraints on Chemical Species 102

Table 5.3. Optimized Relative Reaction Rates for Carbon Limited Penicillin Production 107

Table 5.4. Optimized Relative Reaction Rates for Energy and Reduction Limited Penicillin Production 112

Table 6.1. Reaction Constraints 132

Table 6.2. Constraints on Chemical Species 136

Table 6.3. Optimized Rates for Carbon Limited Production 138

Table 6.4. Optimized Reaction Rates for Reduction Limited Synthesis 141

Table 6.5. Optimum Rates with Acetate as Byproduct 145

Chapter 1. Summary

1.1. The Problem

A vexing issue in working with biochemical systems, especially whole cell systems, is that of working with mass balances. The large number of reactions and the complexity of interrelations of the enzymatic reaction steps which may be present in what is an apparently simple fermentation problem can pose significant difficulty in quantitative analysis of the system. In particular, the structure of the reaction network imposes constraints which normally preclude the use of simple macroscopic mass balances. A common case is that the creation of one desired product will often necessitate the appearance of a second, not necessarily desirable, product.

In a well understood biological system, these constraints may be well known and utilized. A classic example is the predictable production of ethanol and carbon dioxide by yeast grown anaerobically on sugars. Current technology provides the capability to directly introduce, remove, or modify single reactions or entire pathways in microorganisms.

The manipulation and construction of biochemical reaction pathways using the techniques of recombinant DNA technology is also

known as metabolic pathway engineering. Metabolic pathway engineering applications range from improving product yields by modifying existing pathways to constructing entirely new pathways for the synthesis of novel products. The rational design and analysis of these modified pathways requires an ability to understand and predict the effects of changes made to the metabolism of microorganisms.

The determination of theoretical yields for a given pathway and even the derivation of a closed, consistent mass balance on an experiment are commonly confounded by the necessity of working with a large number of linear equations representing the stoichiometry of the reactions in the system. The size of this set of equations effectively precludes the use of manual procedures to work on it. A need then arises to use a machine to work on the problem. This requirement turns an apparently simple operation, working on a mass balance, into a requirement to invest significant effort to get at a basic element of information.

1.2. An Approach to a Solution

In order to remove much of the effort from the process of working with reaction networks, a software package was developed to handle to the mechanical aspects of working on the stoichiometry of biochemical networks. It is hoped that this package will provide a useful design tool

for performing analysis of existing and notional pathways for metabolic pathway engineering. It is relatively simple to use and is flexible enough to find use in variety of applications ranging from biochemical engineering to molecular biology.

The software package, the Pathway Toolbox Program, was constructed consisting of two major components: a database and an optimization routine. The database was geared to the specific purpose of storing and manipulating stoichiometric data with a minimum amount of repetitive effort and a large degree of flexibility. Taking advantage of the extensive amount of work already thoroughly developed in the area of linear optimization, a commercial optimization package, MINOS (Murtaugh and Saunders, 1987), was used to handle the actual numerical aspects of system optimization. The entire software package was designed to run on the Macintosh II computer (Apple Computer Incorporated, Cupertino, California).

The database portion of the software was constructed with a simple to use, graphically flexible language, HyperCard (Apple Computer, Inc., 1988). While some speed and power is lost in using this language, the time needed to write and troubleshoot the program is vastly reduced in comparison to writing a similar program in a more conventional programming language.

Three primary functions are performed by the database. It stores data on chemical species and reactions. It also provides the user a reliable and fairly simple method to manipulate that data. And, finally, it serves as the link, a friendly front end, between the user and the optimization package.

Data storage is relatively flexible. Formulas for species and stoichiometry for reactions are the primary elements of information entered on each record in the database. Additionally, graphics, such as drawn or scanned images of structures, can be stored. Free form text data, such as notes or bibliographic references are also maintained for each data record. The records for chemical species automatically calculate molecular weights and number of available electrons (used for a shortcut mass balancing method) for each species. The records for reactions can check that the posited reaction is elementally balanced.

Data manipulation in the database is aimed toward building standard Mathematical Programming System (MPS) formatted files which describe a particular optimization problem and can be read by optimization packages. To do this, the database provides a mechanism for assembling the problem and setting constraints on it in a relatively user friendly and intuitive manner.

Serving as a front end to the optimization routine, the database then performs the mechanical effort of correctly formatting the problem, writing necessary files, and calling the optimization package to work on the problem. The database also takes the results files from the optimization routine, reads them, and presents the results in an easily readable format.

The optimization routine used was MINOS 5.1 compiled to run on the Macintosh II. The full MINOS package was not used. A streamlined set of routines was compiled specifically to handle the linear optimization problems in reaction network stoichiometry.

1.3. Some Case Studies

Three case studies are presented here as an overview of what might be learned from the analysis of the stoichiometries of biochemical reaction networks. Analyses were conducted on the pathways for the biosynthetic production of astaxanthin, penicillin, and 1,3-propanediol.

Astaxanthin is a forty carbon molecule which is a natural red pigment. It serves as the pigment in a number of organisms and has potential importance as a feed additive for commercial animal husbandry. The analysis conducted here was to determine what the

potential yield of astaxanthin might be when produced by a yeast grown on a glucose feedstock under several different conditions.

A common antibiotic for more than forty years, penicillin is commercially produced by fermentation. However, analysis to determine the maximal yields of penicillin was not done until the late 1970's. This may in part have been due to the large number of reactions which are involved in penicillin synthesis. An analysis of this process was conducted here to check the linear optimization approach to analyzing the pathway against an accepted result. A good agreement was obtained for the expected optimal yields. The optimization approach also provided information on the expected relative reaction rates for optimal yield.

The production of 1,3-propanediol (1,3-PD) presents an opportunity to look at an artificial system and gain some insight into what might be important in creating such a system. 1,3-PD is produced by *Klebsiella pneumoniae* naturally. The enzymatic steps for the production of 1,3-PD from glycerol are small in number and relatively well understood. Other aspects of the metabolism of *Klebsiella pneumoniae* are not so well understood. The metabolism of *Escherichia coli* on the other hand is quite well studied, but *E. coli* does not produce 1,3-PD. Introducing the 1,3-PD pathway into *E. coli* would provide the capability to produce 1,3-PD in a system which is well understood and might consequently be more readily controlled or modified. An analysis is conducted here to

determine what yields of 1,3-PD might be possible under some differing conditions. The analysis provides insight into what metabolic pathways must be expressed for maximum yield and what limits are present in the absence of an important pathway.

1.4. Conclusions

The work done here demonstrates that linear programming techniques can be applied to problems involving the stoichiometry of biochemical reaction networks. Automation provides a means to reduce the amount of effort and increase the accuracy involved in this process. The information provided by the optimization process includes the net reactions of the optimized systems. Information is also provided on the relative rates of the various enzymatic reactions which are needed to produce the optimum yield.

Chapter 2. Literature Review

2.1. Biochemical Modeling

The general problem of modeling biochemical reaction systems has been extensively examined from a number of different angles. At one extreme is the use of gross macroscopic mass balances which do not examine the structure of the biochemical reaction network. At a far greater level of complexity, artificial intelligence methods are employed to build complex models capable of making approximations to allow kinetic modeling of reaction systems about which insufficient data exists to produce a complete dynamic model from experimental data.

The complexity in these models may be considered in several levels. The simplest level is order of magnitude analysis of the overall net reaction taking place in the system without regard to its constituent reactions or kinetics. The second level of effort is in constructing a structured stoichiometric model which attempts to accurately reflect the stoichiometry of the individual reactions in the network. At the third level of model complexity, thermodynamic aspects of the behavior of the reaction system are modeled. The final level of model complexity involves kinetic analysis of the system to capture the dynamic behavior of the system.

This work has centered on a general approach to building and manipulating stoichiometric models of biochemical systems. A driving force behind working with stoichiometric models is the paucity of kinetic and thermodynamic data for many of the biochemical systems which may be of interest (Seressiotis and Bailey, 1988). While individual enzymatic reaction steps in a process may be known directly or inferred from comparison with analogous processes, data on the kinetics of those steps are often not readily available nor are they readily determined for the large number of steps and variety of reaction conditions which may be of interest.

2.2. Detailed Kinetic Models

A few detailed models of the dynamic behavior of biochemical systems have been constructed. These include models of *Escherichia coli* by Domach et al.(1984), Shu and Shuler (1989), and Steinmeyer and Shuler (1989), as well as a model of human red blood cell metabolic dynamics (Palsson et al, 1987; Joshi and Palsson, 1988; Joshi, 1988). The quantity of data required for the construction of dynamic models is quite large. Kinetic information on several dozen reactions and concentration data on approximately the same number of chemical species is required to get a close approximation of real behavior. Mavrovouniotis et al.(1988, 1989) have applied techniques of artificial

intelligence to permit computer modeling of the approximations sometimes used to work with kinetics of enzymatic reactions when compete rigorous data on the kinetics of the reaction are not available. This procedure attempts to capture the heuristic rules commonly used to get order of magnitude estimations of dynamic behavior. However, the number of detailed kinetic models and the amount of information required to build them is still quite small in comparison to the number of biological systems studied.

2.3. Early Simple Models

Since the early days of biochemical experimentation, mass balancing has sometimes been used to help clarify the precise nature of biological reaction networks. An early example of the use of elemental balances in biochemistry is the work of Hoover and Allison (1940) in examining aspects of *Rhizobium melioti* metabolism. They developed a simple model consisting of three equations which could be linearly combined to account for the different respiration quotients (amount of CO_2 evolved per amount of O_2 consumed) observed in rhizobia cultures grown with two different nitrogen sources, ammonia and nitrate. While the reactions proposed represented only macroscopic mass balances, not the individual enzymatic reaction steps, they do have some physical significance in that they each represented the result of an overall

metabolic process such as incorporation of carbon into cell mass and reduction of nitrogen to be incorporated into cell mass.

2.4. Pathway Determination

More recently, a number of researchers in the field of biochemical engineering have made use of macroscopic mass balances and stoichiometric models, or mass balances based on individual reaction stoichiometry, in a variety of applications to the growth of microorganisms. In this discussion, the term fermentation is used to refer to the general process of growing a culture of organisms. The more restrictive sense of fermentation, growing a culture without an exogenous source of oxidizing power, is not used in this chapter.

Verhoff and Spradlin (1976) used a knowledge of the reaction stoichiometry to determine the biochemical pathway for citrate production from glucose by *Aspergillus niger* . Starting with the assumption that the organism made use of commonly known enzymatic reactions, they used these reactions to outline several possible biochemical pathways for citrate production. Then employing a knowledge of certain characteristics of the actual production of citrate, such as the observation that oxalate is produced by the organism when it produces citrate, they worked out which pathways *Aspergillus* might actually use in citrate production. The solutions were determined by

considering linear combinations of reactions which would give rise to observed results. The unknown variables were the relative net rates of each reaction over the course of citrate production. The set of 11 reaction rates was solved subject to the constraints of the observed net production and consumption of the 10 chemical species directly involved in the citrate production reactions. The purpose of their work was to gain a better understanding of what factors might be important in commercial citrate production processes.

Aiba and Matsuoka (1979) also looked at citrate production, though in their work production was by *Candida lipolytica* . They used a stoichiometric approach similar to Verhoff and Spradlin to analyze the production pathways, but followed this up with experimental confirmation. Initially they found three possible pathways that would give rise to the experimentally observed product yields. An initial examination of these pathways revealed that in two of them reactions would be required to occur in the reverse of the normally expected direction, thus casting some doubt as to whether they would actually operate as proposed *in vivo*. Assays for enzyme activity of a few enzymes in the organism revealed *in vitro* activities below the *in vivo* activities which would be expected from the same two possible pathways. Since *in vitro* enzyme activity is assumed to represent a maximal possible reaction rate, not exceeded *in vivo*, the assays provided further evidence for a single possible pathway for citrate production.

Reardon et al. (1987) have applied reaction stoichiometry to identifying metabolic pathway reaction rates. Using the method proposed by Papoutsakis (1984) to build a fermentation equation for *Clostridium acetobutylicum* , they then analyze experimental product formation data to determine the specific reaction rates between branch points in the pathway. As a result of this they are able to determine the direction of several enzymatic reactions and to better understand the functioning of the pathway in fermentation. Of particular interest, the authors were able to show from this analysis that one particular enzyme, NADH:ferrodoxin oxidoreductase, operates in different directions over the time of the fermentation. This observation was supported by other measurements conducted during the fermentation. The information on the enzymatic reaction direction change was important in understanding the mechanism used by the cell to maintain its balance of oxidizing and reducing species.

2.5. Yield Analysis

Cooney and Acevedo (1977) made use of the stoichiometry of individual enzymatic reactions in calculating the maximum theoretical yield for the production of penicillin from glucose by *Penicillium chrysogenum* . Working from a set of observed and generally accepted individual reactions in penicillin biosynthesis, the authors derived a

general stoichiometric equation for the overall process. They were then able to conduct analyses in a relatively simple manner for the effects of various parameters and assumptions on the overall process efficiency. From the analysis, they were able to show that even high yield penicillin fermentations only achieve ten percent of the theoretically possible yield for conversion of glucose to penicillin. Cooney (1979) demonstrated that if the penicillin yield were to be increased, there would be other benefits for the production process in increased volumetric productivity and a reduced oxygen demand by the cells.

The above stoichiometric approach is somewhat clearer and simpler to understand than a similar, more classical biochemical analysis of penicillin production by Heijnen et al.(1979). The latter study made use of elemental balances on the substrates and products in conjunction with macroscopic mass balances to analyze the penicillin fermentation process. The details of whatever biochemistry occurs inside the cell is contained in a number of yield parameters, Y_{ij}, the amount of substrate i produced per substrate j consumed. The results indicate a maximum yield of penicillin of 0.46 mole penicillin per mole glucose, while Cooney and Acevedo derived a maximum yield of 0.56 mole penicillin per mole glucose for the same conditions. While both are far above observed yields, the results of Cooney and Acevedo provide some mechanistic insight into the process, while the approach of Heijnen et al.

leaves the mechanism hidden in the macroscopic parameters of yields of one chemical species on another species.

Stouthamer has made an extensive study of mass balances in biochemical processes. This includes work on both elemental macroscopic balances and balances based on the stoichiometry of individual reactions. An analysis (Stouthamer, 1973) of the energy requirements for cell biosynthesis, in terms of adenosine triphosphate (ATP) requirements, was based on a linear combination of the requirements for the synthesis of the various cell components. In turn, these were derived from the known, or accepted, energy requirements for the reaction steps in the component synthetic pathways. From this essentially stoichiometric analysis, Stouthamer determined the maximum possible yield of cell mass from ATP for a number of media compositions, and showed how experimentally observed yields could be explained by simple linear combinations of these maximum yields and changes in cell growth rates and maintenance energy requirements. A later study (Stouthamer, 1977) expanded this effort to look at the effects of differing nitrogen sources for cell growth along with differing energetic yields of ATP from the reduction of those nitrogen sources. In this effort, relatively good agreement was found between what was predicted from the theoretical analysis and what was observed in culture growth.

Roels (1980, 1983) has done extensive work in establishing general procedures for using macroscopic elemental balances, balances based on reaction stoichiometry, and energy balances to problems in biotechnology. While some of this work is similar to that of Aris (1969), it is expressed in terms of biochemical systems, and addresses some aspects specific to fermentation processes. These areas include the problems associated with changing cell densities and changing yields of products with time in fermentations.

2.6. Stoichiometry in Fermentation Control

Cooney, Wang and Wang (Cooney et al.,1977; Wang et al., 1977, 1979) have proposed and demonstrated some methods of controlling fermentations to increase volumetric productivity and product conversion yields based on the overall reaction stoichiometry. By monitoring the time course of the fermentation in terms of the overall reaction stoichiometry, they were able to control the fermentation and increase productivity without needing to estimate or assume kinetic parameters or yield coefficients for the growth of bakers' yeast, *Saccharomyces cerevisiae* . The effort was necessary because important variables in the fermentation, such as growth substrate and cell mass concentrations, were not directly measurable in a time scale appropriate for control use. From mass conservation requirements for the elements present in the reactants and products, important unmeasured variables

were determined indirectly from measurements made on more readily measured variables. Feeds were controlled to minimize inhibition of cell growth and of formation of ethanol, an undesirable byproduct.

Swartz and Cooney (1979) used the same method to control the production of single cell protein from *Hansenula polymorpha* grown on methanol. Here the control problem is somewhat more involved than that for the bakers' yeast fermentation, because high concentrations of the methanol substrate inhibit cell growth, and accumulation of two major intermediate species of methanol metabolism, formaldehyde and formic acid, both inhibit growth and decrease the quality of the end-product single cell protein. In this process, as the authors put it, "continuous monitoring of the stoichiometry of cell formation" to control substrate addition improved the process results. Byproduct formation was reduced while volumetric productivity and product conversion yield were increased.

2.7. Stoichiometry and the Fermentation Equation

Papoutsakis (1984, 1985 a, b) has proposed a method to formulate what he terms a fermentation equation for a given biochemical reaction system. Starting with a knowledge of, or assumptions about, what individual enzymatic reactions take place within a given organism, one builds a system of stoichiometric equations describing these reactions.

After assigning a reaction rate variable to each reaction, a linear combination of the stoichiometric equations multiplied by their rate variables yields an overall stoichiometric equation describing the net reaction of the fermentation. With some knowledge of reactant and product initial and final concentrations, as well as some steady state assumptions for intracellular intermediate species, one can solve the resultant set of equations to check for consistency in experimental results. The equation can also be used to predict maximum yields for conversion of substrates to desired products through a given biochemical pathway or set of pathways. Papoutsakis gives several examples of the use of this technique in analyzing results from the fermentative production of a number of organic acids and alcohols. Apparently, Papoutsakis uses a trial and error method of solving the equations in his articles, though he does propose that standard linear programming routines might be used for examples more complex than the ones he presents.

Tsai and Lee (1988) have recently proposed a rapid method for the analysis of microbial metabolism based on mass balances and, in some cases, reaction stoichiometry. They examined the problem of how much of the stoichiometric structure of a biochemical pathway must be used in the analysis of the overall effect of the pathway, that is, whether a linearly independent set of reactions from the pathway was sufficient to fully describe the overall reaction of the pathway. Applying Gibbs' rule to

a biochemical reaction network of R independent reactions involving S chemical species, they show that data from the network structure is needed only if:

$$R - S - 2 > 0 \qquad (2.1)$$

If the value is zero, then the network may be analyzed by using the set of R independent reactions alone. A difficulty with this method is that there is still a fair amount of work involved in finding a set of independent reactions for a large biochemical reaction network, and more work yet in quickly discerning what the appropriate additional constraints should be if they are needed.

2.8. Stoichiometric Data Analysis

Wang and Stephanopoulos (1983) have used elemental mass balances based on overall reaction stoichiometry as a basis for the statistical analysis of fermentation measurements. They extend the procedures of Minkevich et al.(1973) and Cooney et al.(1977) by using them as the basis for a statistical analysis of fermentation data. Through the analysis of the results, rather than simply averaging errors out over a number of sources, it is possible to isolate the source of the error. In a real time analysis of a fermentation, this can provide an indication of an equipment malfunction or an upset to the biological system.

Stephanopoulos and San (1984) have proposed a more elaborate use of the general procedures outlined by Wang et al.(1977) to determine fermentation variables. Starting with the general elemental balances for reactants and products in the fermentation process, Stephanopoulos and San propose procedures to determine a wide range of fermentation culture properties and variables from available measurements. They propose methods for handling transient conditions for cultures undergoing dynamic changes in environment, as well as for processing sensor signals to smooth out the noise expected from process measurements.

Erickson et al.(1978 a, b) have proposed a general approach to working with the macroscopic mass balances of biochemical reaction systems in order to analyze fermentation results. Making use of the relatively constant values of certain aspects of cell mass balances, they shorten the amount of calculational effort needed to perform mass and energy balance calculations on biological systems. One of the relatively constant values in cells is the reductance degree of cell mass. The reductance degree, γ, of a typical biological compound, $CH_hO_yN_n$, is defined by Minkevich and Eroshin (1973) as:

$$\gamma = 4 + h - 2y - 3n \qquad (2.2)$$

This is based on normalizing the compound molecular formula to a single atom of carbon per molecule. Actually, the degree of reductance serves as a shorthand method to assist in rapidly checking the

completeness of the overall mass balance for a reaction. A carbon balance and a reductance degree balance serve as a quick mass balance for a typical biochemical reaction. Erickson demonstrates simple ways to use these methods to check for consistency of data from fermentations, as well as to reveal the presence of undetected products from the fermentation.

2.9. Other Uses of Stoichiometry in a Biochemical Network

Irvine and Bryers (1985) made use of a mass based, rather than the more common molecular based, stoichiometry in the analysis of biochemical reaction networks in waste treatment processes. The mass basis was used because of the relative simplicity of determining overall effluent and sludge compositions as opposed to rigorously identifying the elemental composition and relative amounts of a myriad of chemical species present in waste treatment operations. They then extended the methods of Aris (1969) in handling molecular based reaction stoichiometry to analyze limitations and yields in waste treatment on a mass formula basis.

Seressiotis and Bailey (1986, 1988 a, b) have used artificial intelligence to work with reaction stoichiometry. Their approach has focused on deriving potential pathways for production of a given product from a designated substrate. The program they developed, Metabolic

Pathway Synthesis (MPS), uses a database of information on enzymatically catalyzed reactions to assemble and propose possible pathways for the desired conversion. The program output lists the enzymes which would participate in the proposed pathway along with the net reaction the pathway would catalyze.

Mistry and Cooney (1989) have done some stoichiometrically based analysis of ethanol production by *Clostridium thermosaccharolyticum* . Using the fermentation equation approach of Papoutsakis, they developed a linear model of the pathways giving rise to the major products found in fermentations of this organism. They then elucidated what the possible product distributions would be by considering varying product ratios at the branch points in the pathway. Their goal was to optimize ethanol production from a xylose feed. From the knowledge of how the overall pathway should be operating to get a maximum yield, it should be possible to manipulate fermentation conditions to induce the pathway to function at this optimal level. This was possible because the modification of fermentation conditions could change product distributions in manner which could be explained by an altered flux through certain portions of the pathway. Mistry and Cooney found that the predicted product distributions did correlate well with experimentally observed product yields.

Chapter 3. Linear Modeling of Reaction Networks

3.1. Quantitative Analysis of Biochemical Systems

A problem in most aspects of quantitative analysis of a chemical transformations carried out by biochemical systems, especially those produced using whole cells as the catalyst, is the large number of chemical species and enzymatic reaction steps involved. Even such a basic step in understanding the process as conducting a detailed mass balance over the reaction system is made difficult because of the large quantity of information to be handled. As Kootsey (1987) has pointed out, biochemists often conduct analyses of some aspects of the systems through intuitive hypothesis testing rather than wrestling with the mathematics needed to get a fuller quantitative understanding of the process.

3.2. The General Approach to Stoichiometric Notation

A method of mathematically representing the stoichiometric information of reaction networks was well developed by Aris (1963, 1969) in the use of matrix notation common to linear algebra. The details of this notation are thoroughly presented by several authors (Aris, 1969; Roels, 1983; Sørenson and Stewart, 1980) and here only a brief summary

appropriate to the stoichiometric model effort is presented to establish a standard notation for this work.

In a reaction system which consists of a number of reactions involving several chemical species, each reaction may be represented by a reaction notation:

$$\nu_{11} S_1 + \nu_{12} S_2 = \nu_{13} S_3 + \nu_{14} S_4 \qquad (3.1)$$

$$\nu_{21} S_1 + \nu_{23} S_3 = \nu_{25} S_5 \qquad (3.2)$$

where ν_{ij} is the stoichiometric coefficient of species j in reaction i. Using the standard convention of giving stoichiometric coefficients of reaction products positive values and those of reactants negative values, the same reactions may be represented as:

$$-\nu_{11} S_1 - \nu_{12} S_2 + \nu_{13} S_3 + \nu_{14} S_4 = 0 \qquad (3.3)$$

$$-\nu_{21} S_1 - \nu_{23} S_3 + \nu_{25} S_5 = 0 \qquad (3.4)$$

The reactions can be more compactly represented for simpler handling by defining a stoichiometric matrix, **S**. In matrix **S** each row corresponds to a reaction and each column to a chemical species. Thus each element ν_{ij} of **S** represents the stoichiometric coefficient of species j in reaction i. For example the stoichiometric matrix for the two reactions described above is:

$$\begin{bmatrix} -\nu_{11} & -\nu_{12} & \nu_{13} & \nu_{14} & 0 \\ -\nu_{21} & 0 & -\nu_{23} & 0 & \nu_{25} \end{bmatrix}$$

Continuing with this simple example, the rate of the first reaction, v_1, and that of the second, v_2, may be represented in a column vector **v**:

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

The rate of production of each chemical species, R_j is then given in the column vector, **R**, which is the product of the transpose of **S** with **v**:

$$\begin{bmatrix} -\nu_{11} & -\nu_{12} & \nu_{13} & \nu_{14} & 0 \\ -\nu_{21} & 0 & -\nu_{23} & 0 & \nu_{25} \end{bmatrix}^T \bullet \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} R_{S1} \\ R_{S2} \\ R_{S3} \\ R_{S4} \\ R_{S5} \end{bmatrix} \tag{3.5}$$

This may be more succinctly expressed as:

$$\mathbf{S}^T\mathbf{v} = \mathbf{R} \tag{3.6}$$

3.3. Biochemical Systems as Open Systems

One difficulty with many biochemical reaction systems, for instance fermentations, is that there is little information on the individual reaction rates. The needed data on intracellular metabolite

concentrations and reaction rates is usually insufficient to allow derivation of a reaction rate expression which will be valid over the potentially wide range of conditions to be modeled.

One approach to this problem is to analyze the fermentation as an open system with the reaction rates as a set of unknown variables. In this method of analysis the cell mass in the fermentation is considered as an open system with substrates flowing in and products, and in some cases additional cell mass, flowing out. The rates of the various reactions in the system are assumed to be independent of time and can be handled as unknown numerical constants. While this method does not allow for the exploration of the dynamic aspects of the system, it does provide a rapid way to examine the different possible product distributions which the system may generate.

Commonly, researchers have analyzed the stoichiometry of the biological reaction system by explicitly writing out the equations for the rates of production of each species in the system (Papoutsakis fermentation equation,1984; Tsai's rapid method, 1988). For example, the rate of production of species S_3, R_{S3} , in the simple example above would be:

$$R_{S3} = \nu_{13} v_1 - \nu_{23} v_2 \qquad (3.7)$$

With some assumptions about cellular intermediates (to be discussed further) and with some knowledge about the potential feed and product

streams, the unknown reaction rates may, in some cases, be determined. With these reaction rates, the the rates of production of species which may not have been directly set may be determined. Similarly, experimental fermentation data may be analyzed for internal consistency.

3.4. Linear Programming

The manual approach described above is actually an explicit method of obtaining the solution to a linear programming problem or linear optimization. This solution can be more easily and more generally obtained using the techniques developed specifically to solve linear programs. The details of the handling and solution of these problems is well developed (Fletcher, 1987; Noble and Daniel, 1988; Strang, 1980; Sedgewick, 1984) and is discussed here briefly only in relation to the fermentation problem.

An organism grown in culture carries out some set of biochemical reactions which can be used to develop the general approach. The net overall reaction at any instant in time over the course of a fermentation of the organism may be described by equation 3.6:

$$\mathbf{S}^T\mathbf{v} = \mathbf{R} \qquad (3.6)$$

If an attempt is made to find a limiting value for a specific element of **R**, such as the production of a high value chemical species, by solving the

matrix equation, what is actually undertaken is a linear optimization problem. The solution will depend on the constraints on the possible values of elements of **R** and **v**, not simply on the elements of **S**. The manual solution of this problem is dauntingly difficult for all but the very simplest systems. But if the problem is understood in more general terms, many automated procedures are available to handle it.

If some bounds are placed on the possible values which the elements of R may possess (these bounds will be discussed further below) then equation 3.6 may be expressed more generally as:

$$\mathbf{S}^T \mathbf{v} \leq \mathbf{R} \tag{3.8}$$

Equation 3.8 is one case of a general linear equation:

$$\mathbf{S}^T \mathbf{v} + \mathbf{I}\mathbf{s} = \mathbf{R} \tag{3.9}$$

Here an additional vector is added consisting of the column vector product of an identity matrix, **I**, and a matrix of slack variables, **s**.

3.5. Manipulating Inequalities

Slack variables allow the efficient handling of inequalities by treating them as strict equalities involving slack variables. For instance, the inequality with variables x_1 and x_2:

$$a\,x_1 + b\,x_2 \geq 0 \tag{3.10}$$

may be written as an equality:

$$a\,x_1 + b\,x_2 + s_1 = 0 \tag{3.11}$$

where s_1 is a slack variable subject to the limits:

$$-\infty \leq s_1 \leq 0 \qquad (3.12)$$

Similarly, each original variable, x_1 or x_2, may be subject to some upper and lower bounds on its values.

Thus if some potential limits on the possible values of the reaction rates (to be addressed further) in the fermentation are considered, the linear problem of equation 3.9 may be restated as:

$$\mathbf{S}^T \mathbf{v} + \mathbf{I}\mathbf{s} = 0 \qquad \mathbf{L_v} \leq \mathbf{v} \leq \mathbf{U_v} \quad \mathbf{L_s} \leq \mathbf{s} \leq \mathbf{U_s} \qquad (3.13)$$

with $\mathbf{L_v}$ and $\mathbf{L_s}$ as the column vector containing the lower bounds for the reaction rates and slack variables respectively and $\mathbf{U_v}$ and $\mathbf{U_s}$ as the upper bounds.

The nature of the sets of bounds should be clarified. The limits on the slack variables can be seen directly from the inequality which gave rise to the variable. For instance in the example of the inequality:

$$a\,x_1 + b\,x_2 \leq 0 \qquad (3.10)$$

becoming the equality:

$$a\,x_1 + b\,x_2 + s_1 = 0 \qquad (3.11)$$

it is clear that s_1 must be non-positive. That is

$$-\infty \leq s_1 \leq 0 \qquad (3.12)$$

Analogously, for the inequality:

$$c\,x_3 + d\,x_4 \leq m \qquad (3.14)$$

becoming to the equality:

$$c\,x_3 + d\,x_4 + s_2 = m \qquad (3.15)$$

s_2 must have non-negative values or

$$0 \leq s_2 \leq \infty \qquad (3.16)$$

3.6. Constraints for Reaction Types

The case of the possible limits for the reaction rates is similarly straightforward. For the reaction:

$$\nu_{11}\,S_1 + \nu_{12}\,S_2 = \nu_{13}\,S_3 + \nu_{14}\,S_4 \qquad (3.1)$$

with reaction rate v_1. A positive values of v_1 indicates the reaction occurs in the direction written, that is reactants S_1 and S_2 going to products S_3 and S_4. A negative value of v_1 indicates that the reverse of the written reaction occurs, that is S_3 and S_4 reacting to produce S_1 and S_2. In the case of setting limits for the reaction rates it is sufficient to consider whether or not the reaction is reversible under the conditions of the fermentation. For a reversible reaction the limits for its rate, r_r, are set as:

$$-\infty \leq r_r \leq \infty \qquad (3.17)$$

For an irreversible reaction with rate r_i, assuming that it has been written in the direction in which it occurs, the limits for its rate are:

$$0 \leq r_i \leq \infty \qquad (3.18)$$

3.7. Constraints on Chemical Species

As described in section 3.5, in the case of the overall equation for a fermentation, a more appropriate expression of the general problem is:

$$\mathbf{S}^{T}\mathbf{v} + \mathbf{I}\mathbf{s} = \mathbf{C} \qquad \mathbf{L}_{\mathbf{v}} \leq \mathbf{v} \leq \mathbf{U}_{\mathbf{v}} \quad \mathbf{L}_{\mathbf{s}} \leq \mathbf{s} \leq \mathbf{U}_{\mathbf{s}} \qquad (3.19)$$

where the column vector **C** contains specific values of constraints for the chemical species involved in the reaction network being analyzed.

Since the solution to the linear programming problem yields a set of instantaneous reaction rates for the fermentation system, the elements of the column vector **C** represent the net rates of production of the various chemical species in the system. However, as these production rates also represent the stoichiometric coefficients for the various chemical species in an overall fermentation reaction, the vector **C** can also be considered to be the vector of stoichiometric coefficients for the overall reaction. Thus a product of the fermentation will have a positive coefficient in **C**, and a substrate in the fermentation will have a negative coefficient in **C**. A species which is neither a product nor a substrate in the overall reaction will have a coefficient of zero, though it may indeed participate as an intermediate species for which there is neither a net consumption nor a net production.

Mavrovouniotis (1988) gives a detailed discussion of what the constraints in **C** may be for a variety of cases. He derived the constraints

in connection with the problem of enumerating all the possible metabolic paths for the conversion of one species into another, but the constraints are equally applicable to the of linear optimization of a metabolic network. For instance, in a carbon-limited fermentation which starts with a fermentation broth containing 100 moles of glucose as the carbon source and produces a product such as an organic acid or alcohol, the net consumption of glucose will not be more than 100 moles. A statement of the constraints on the net glucose production, R_G, for the fermentation system is:

$$-100 \leq R_G \leq 0 \qquad (3.20)$$

Similar general constraints for the overall reaction coefficient R_{Si}, the net production of a given chemical species S_i, are given in Table 3.1 (adapted from Mavrovouniotis, 1988).

Table 3.1. Constraints on Chemical Species Coefficients, R_{Si}

(adapted from Mavrovouniotis, 1988)

S_i is a reactant but not a product:	$-\infty \leq R_{Si} \leq 0$
S_i is a reactant with coefficient not greater than c_i	$-c_i \leq R_{Si} \leq 0$
S_i is a reactant with a coefficient of at least c_i	$-\infty \leq R_{Si} \leq -c_i$
S_i is a reactant with a specified coefficient c_i	$R_{Si} = -c_i$
S_i is a product but not a reactant	$0 \leq R_{Si} \leq \infty$
S_i is a product with a coefficient not greater than c_i	$0 \leq R_{Si} \leq c_i$
S_i is a product with a coefficient of at least c_i	$c_i \leq R_{Si} \leq \infty$
S_i is a product with a specified coefficient c_i	$R_{Si} = c_i$
S_i is neither a reactant nor a product	$R_{Si} = 0$
S_i is under no a priori constraint	$-\infty \leq R_{Si} \leq \infty$

The numerical values for the c_i's may derived from actual or expected composition of the starting fermentation medium and from analysis of the the actual or expected products in the medium. The c_i's must all be expressed on the same basis as the reaction stoichiometry. Throughout this work, a mole basis is used. In some applications, such as waste treatment processes, a mass basis may be simpler to use owing to the complexity of the medium involved (Irvine and Bryers, 1985).

Most species within a biochemical reaction network will be cellular intermediates which are not expected to accumulate or depleted

during a fermentation. Therefore, the constraint that a species is neither a reactant nor a product can be applied to most species within a biochemical reaction network. For example, acetyl-coenzyme A (acetyl-CoA) would not commonly be seen participating in stoichiometric quantities in an overall reaction representing the fermentation. Additionally, species which serve to carry energy within the cell, such as ATP and ADP, and species which serve as reduction and oxidation (redox) potential intermediates, such as NADH/NAD^+ and $NADPH_2$/NADP, are usually present in the cell in only catalytic quantities. They are not expected to appear in stoichiometric quantities in an overall reaction for the fermentation, especially when considered relative to the concentrations of the major substrates and products. Thus the coefficients of these species can be set to zero in the overall reaction produced by a complete biochemical reaction network.

For a number of chemical species it may prove impractical to determine before the fact whether they will be products, reactants, or whether or not they will be accumulated in the overall reaction. Common examples of these species are protons, H^+, and water, H_2O. For these species, the constraint on their stoichiometric coefficient, R_{Si}, is given as:

$$-\infty \leq R_{Si} \leq \infty \tag{3.21}$$

This is, in fact not a constraint, as the solutions to the equation will always consist of only real numbers. In the linear program, species

having this constraint actually play no role in arriving at a solution to the problem.

3.8. Objective Functions

In some cases equation 3.19,

$$\mathbf{S}^{T}\mathbf{v} + \mathbf{I}\mathbf{s} = \mathbf{C} \qquad \mathbf{L_v} \le \mathbf{v} \le \mathbf{U_v} \quad \mathbf{L_s} \le \mathbf{s} \le \mathbf{U_s} \qquad (3.19)$$

will have no single unique solution. Rather there will be a set of vectors **v** which contain reaction rates satisfying the equations and constraints. The goal of optimization in general is to find the vector of variables, **v**, which gives an optimal, either maximal or minimal, value to an objective function. A typical objective function might be a cost or time function given that each element of **v** has some cost or duration associated with it. Then the overall solution to the problem might be to minimize the cost or time duration of a project or to maximize the monetary profit of a process.

In a fermentation system it is not immediately clear what the objective function is. Whether the cell is attempting to produce a maximal amount of biomass from a substrate or attempting to maximize production of some number of different chemical species is problematic to determine. Indeed, the determination of the form of the objective function for this type of process would also be hard to ascertain. It has been suggested (Ramkrishna, 1983; Turner et al, 1989) that useful

predictions can be made by assuming that the growth rate is maximized under all conditions.

From an engineering standpoint, however, a readily useful objective function can be formulated. Since a fermentation is typically run to produce a single important product, an obvious engineering objective of optimization of the reaction network of the cells in the fermentation would be to maximize the production of this species (Georgiou, 1988). Potentially, a more appropriate objective might be to maximize the profit of a given fermentation by accounting for feed stream component costs and for the values of potential products along with costs associated with downstream processing. In order to concentrate on the general approach of optimizing these systems, the objective used here will simply be the production of a single important chemical species.

From an industrial perspective the chemical machinery in the cell "works" for free; each of the reactions in the cell biochemical reaction network has no monetary cost associated with it, and there is no dollar cost of an optimized reaction network to be calculated or minimized. As kinetic data is often not known beforehand and is, in fact, part of what is derived in the problem solution, minimization of time parameters cannot a part of the objective. Thus the derivation of the un-weighted vector of reaction rates, **v**, giving rise to the maximum (or minimum) value of the

coefficient of the objective species, S_j, in the stoichiometric equation for the overall fermentation process is the initial goal of optimization of the reaction network.

An interesting aspect of the optimization process is that it provides a simple, rapid way for finding a potential biochemical pathway from one metabolite to another if such a path exists. Unlike the method of Seressiotis (1988), this method will not enumerate all the possible pathways connecting two metabolites, but only a pathway which optimizes production of the objective species subject to the imposed constraints.

If the column vector $\mathbf{B}_j$ is defined as the column of the stoichiometric matrix **S** which corresponds to the stoichiometric values of the objective species in the reactions in the network, S_j, then the coefficient of the species to be optimized, R_{Sj}, is given by:

$$\mathbf{B}_j^T \mathbf{v} = R_{Sj} \tag{3.22}$$

The complete linear program problem may then be stated as:

$$\begin{aligned} &\text{maximize} && \mathbf{B}_j^T \mathbf{v} \\ &\text{subject to} && \mathbf{S}^T \mathbf{v} + \mathbf{I}\mathbf{s} = \mathbf{C} \\ &\text{and} && \mathbf{L}_\mathbf{v} \le \mathbf{v} \le \mathbf{U}_\mathbf{v} \quad \mathbf{L}_\mathbf{s} \le \mathbf{s} \le \mathbf{U}_\mathbf{s} \end{aligned} \tag{3.23}$$

3.9. Solving the Problem

Methods for the solution for this type of problem are well developed (Fletcher, 1987; Strang, 1980; Noble and Daniel, 1988; Sedgewick, 1984) and are readily available as standard algorithmic packages such as MINOS (Murtaugh, 1987) and LINDO (Schrage, 1989). Indeed, Noble and Daniel (1988) point out that the safest method of solving linear programs is to use already available software. Given the extensive development and use of these routines there is little to be gained and a considerable amount to be lost in attempting to create one *de novo*.

The actual creation and manipulation of the optimization problem can quickly become problematic even for a relatively small biochemical reaction system if done by hand. The experience of developing this software package indicates that for a system of biochemical reactions the number of chemical species involved is approximately the same as the number of reactions. Thus the stoichiometric matrix **S** for a system of n reactions contains roughly n^2 elements (however, since a typical reaction might involve only two to four species, the matrix is fairly sparse). Relatively simple biochemical reaction networks, such as those examined in further chapters, may involve several tens of reactions, and automated problem formulation and manipulation seems to be the only practical way to analyze these networks. Blum and Stein (1982) have commented on the difficulty in trying to solve by hand accurately even a

moderately sized set of simultaneous equations in a biochemical problem and suggested the use of machine-aided linear algebra to handle them.

3.10. The Pathway Toolbox Program

The overall approach taken here was to construct a software package, the Pathway Toolbox Program, consisting of two components. One of the components handles the storage and textual manipulation of information on reaction networks. The second component is a well-developed and accepted routine, MINOS, used to perform tne calculations involved in linear programming optimization. The general structure of the total software package is shown in Figure 3.1

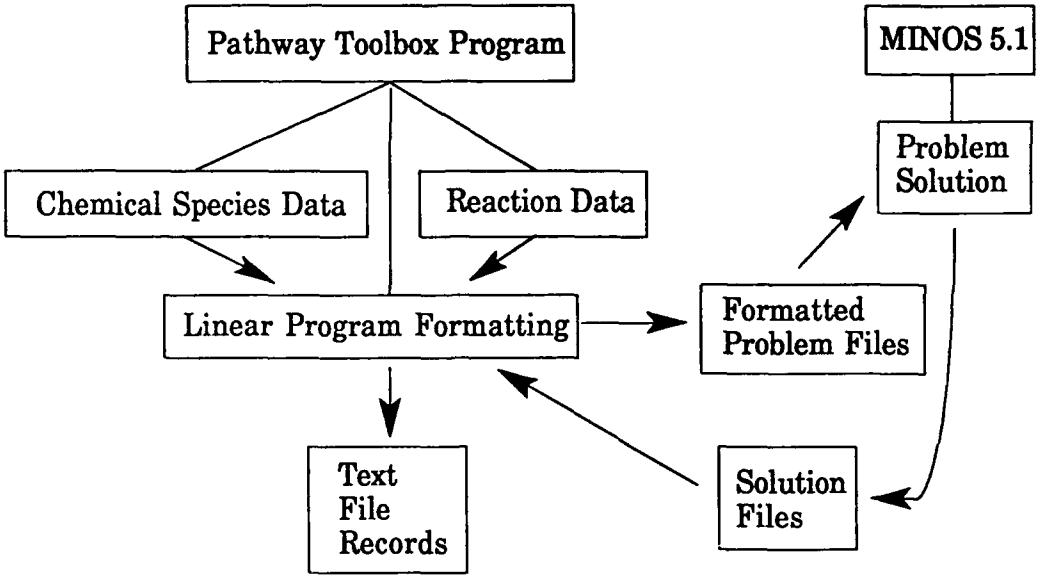


Figure 3.1. Software Structure

The entire software package was developed to be run on a common personal computer so that it could be easily used in experimentation and in teaching. The Macintosh II computer was used to take advantage of its the speed to handle manipulations of the database. MINOS 5.1 was compiled to run on the Macintosh using the Absoft MacFortran/020 Fortran 77 compiler (Absoft, 1989). An abbreviated form of MINOS was used, tailored to handle only linear problems. The remainder of the software package, the Pathway Toolbox Program, was written using Apple Computer's HyperCard. This particular language was chosen because it provides for the rapid and relatively simple creation and

because it provides for the rapid and relatively simple creation and modification of a program with an intuitive, graphically driven user interface.

The Toolbox Program primarily consists of the three components shown in Figure 3.1. One set of records, which in HyperCard are called cards, contains data on individual chemical species. A second set of cards describes the stoichiometry of specific reactions. The third component is a card which, using the reaction and chemical information cards as a basis, forms a linear program reflecting the stoichiometry of the reactions in a reaction system. With this card, the user can establish constraints and designate a chemical species for which production in the system can be optimized. Additionally, this card creates text files describing the problem for MINOS to solve, calls MINOS to solve them, and formats the MINOS output into a readily comprehensible report.

Figure 3.2 is an example of a card for data on a specific chemical species. The user enters the name and chemical formula of the species in the appropriate fields (the boxes shown on the card), and may either import a graphic image of the chemical structure or create one with the HyperCard drawing tools. An additional field, hidden in this example, is available to record other text data. The molecular weight is automatically calculated. The number of available electrons, also called the number of available hydrogens, is calculated using a modification of

the method proposed by Minkevich and Eroshin (1973) for the calculation of γ, the degree of reductance for a biochemical species, $C_cH_hO_oN_nP_pS_sCl_lBr_bF_f$:

$$\gamma = 4c + h + 5p - 2o - 3n - 2s - l - b - f - \text{ionic charge} \quad (3.23)$$

This expression makes some assumptions about the valence of nitrogen being 3, of sulfur being 2, and of phosphorous being 5 in biological systems, but it is sufficient to serve as a consistent basis for determining reduction and oxidation species requirements for reactions. The charge entered is also arbitrarily determined by the assumption the user makes about the ionization of the species in the system.

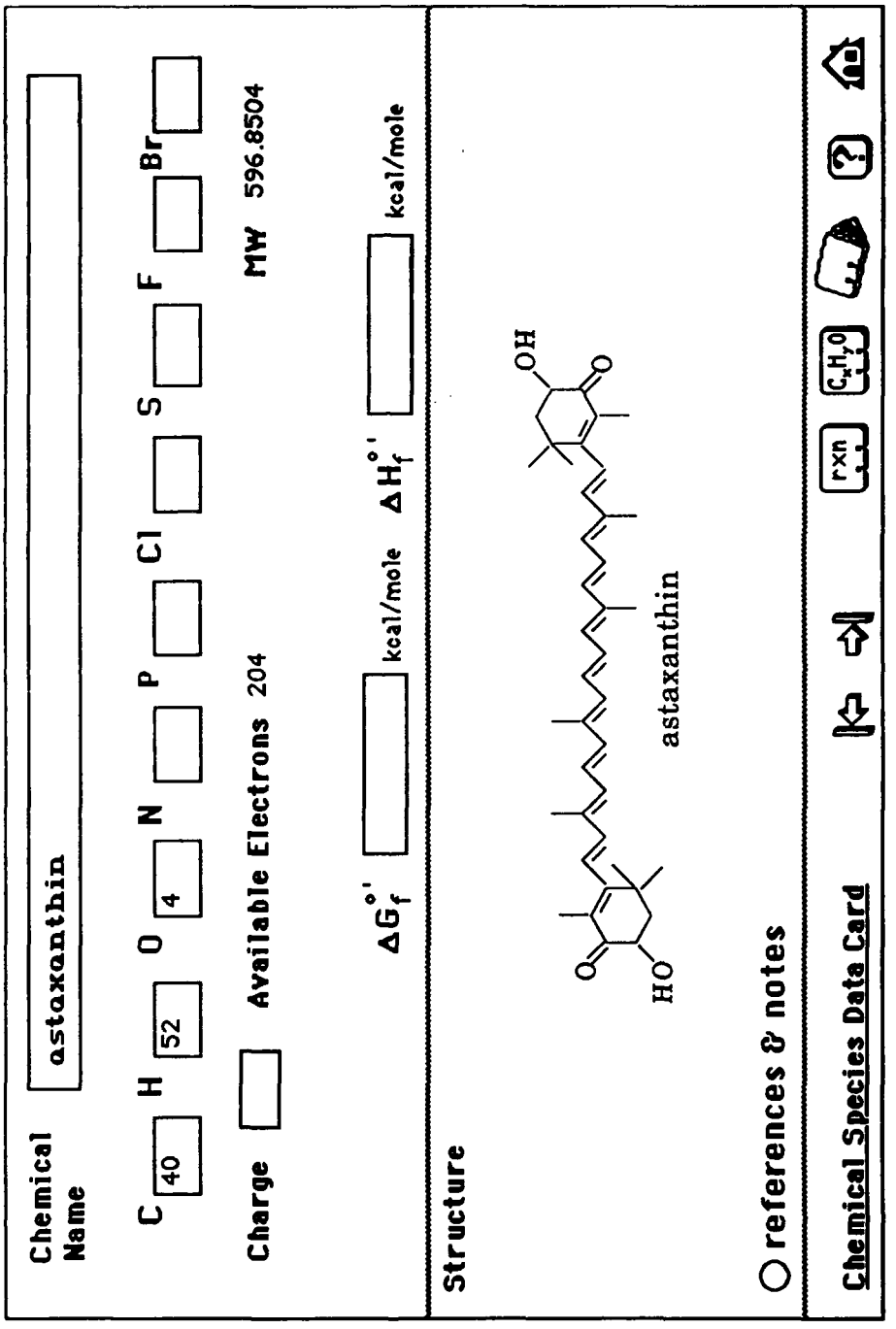


Figure 3.2. Data Card for Information on a Chemical Species from Pathway Toolbox Program.

Figure 3.3 is an example of a data card for a reaction. The various data elements on the card are described below.

Rxn/Enzyme Name: The user enters the name of the reaction in this box.

EC Number: The enzyme code number (International Union of Biochemistry, 1984) may be entered if it is known. Although it plays no role in the problem solution, it does serve as an additional piece of information to describe the reaction.

Irreversible?: The reversibility of the reaction may be designated. Since the specific energy change of the reaction is often unknown, this designation is made by the user based on heuristic rules. This particular piece of data can be altered when setting constraints on a particular problem.

Use Card?: Indicates whether or not the data on this card is to be included in the reaction system to be analyzed.

Substrates/Products: The species involved in the reaction are chosen from a list of species which is generated by the program from the species data cards to avoid computational difficulties arising from misspellings.

The numbers following the species are the stoichiometric coefficients of the chemical species in that reaction.

Balancer: This box can be used to quickly check the elemental balance of the reaction. The check may be useful to discern which species may have been left out of a reaction, since it is not uncommon for published versions of enzymatic reactions in scientific papers and textbooks to not rigorously balance hydrogens and oxygens.

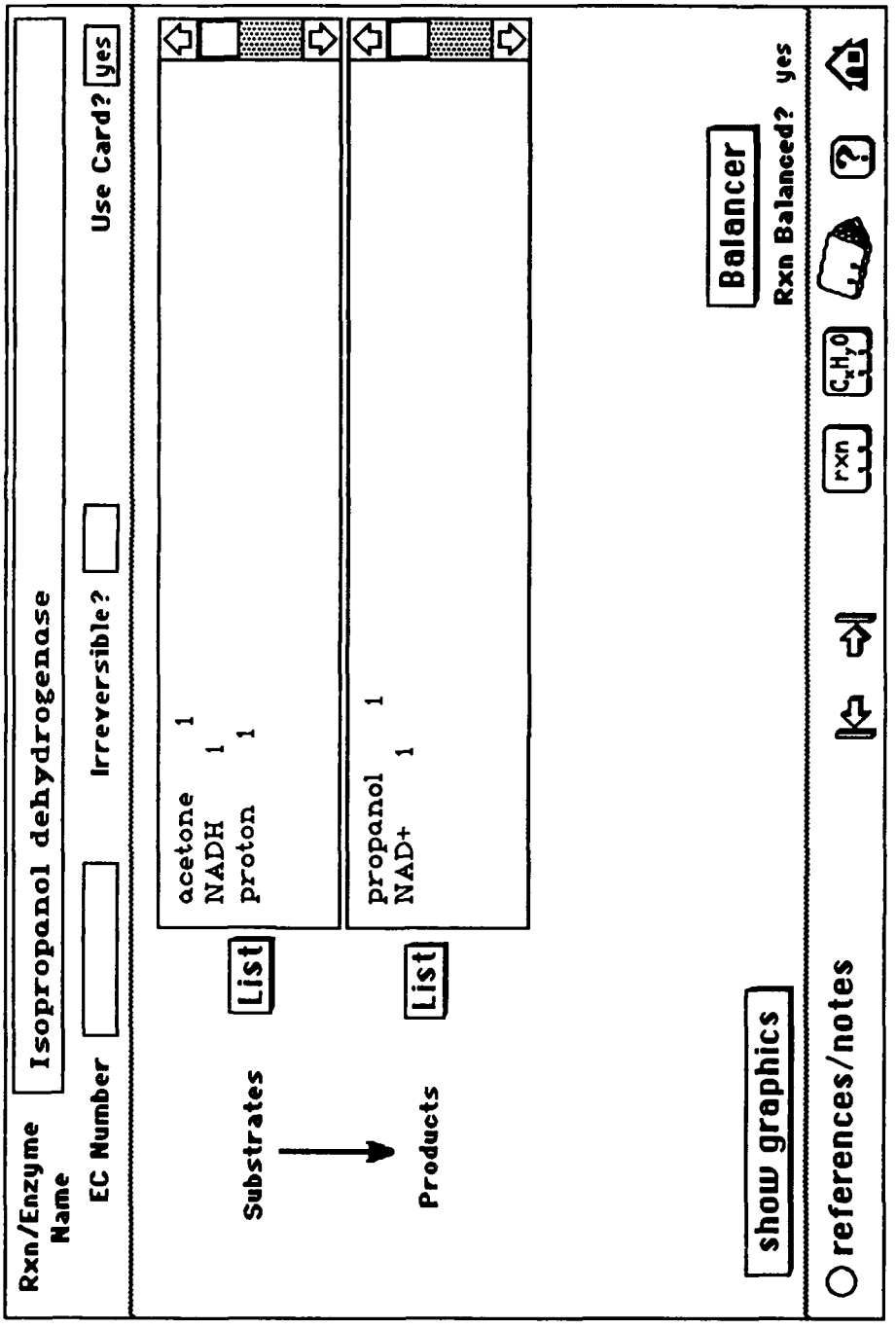


Figure 3.3. Data Card for a Reaction

The card which handles the formulation and manipulation of the optimization problem, the optimization card, is shown in one form in Figure 3.4. This example is of using the card to set the constraints on the coefficients for the chemical species in the reaction system.

The commands for creating a reaction system from the data cards and for manipulating that data are embedded as pop-up menus and graphics objects on the card. Commands are issued using the computer mouse rather than typing on a command line. The "Set Reaction Network" box performs the collating and indexing required to set up the stoichiometric data for a set of reactions. The "Apply Constraints ..." box formats and writes the linear program files and calls the optimization program to operate on those files.

The constraints set may consist of equalities or one-sided inequalities. Constraints to maximize or minimize a value may also be designated. There is also an option to set no constraint on a given species. The default constraint which is set is that a species is not accumulated. This constraint corresponds to the case for most species which are expected to be intracellular intermediates.

The small pictures across the very bottom of all of the different cards are icons which assist the user in quickly navigating through the database.

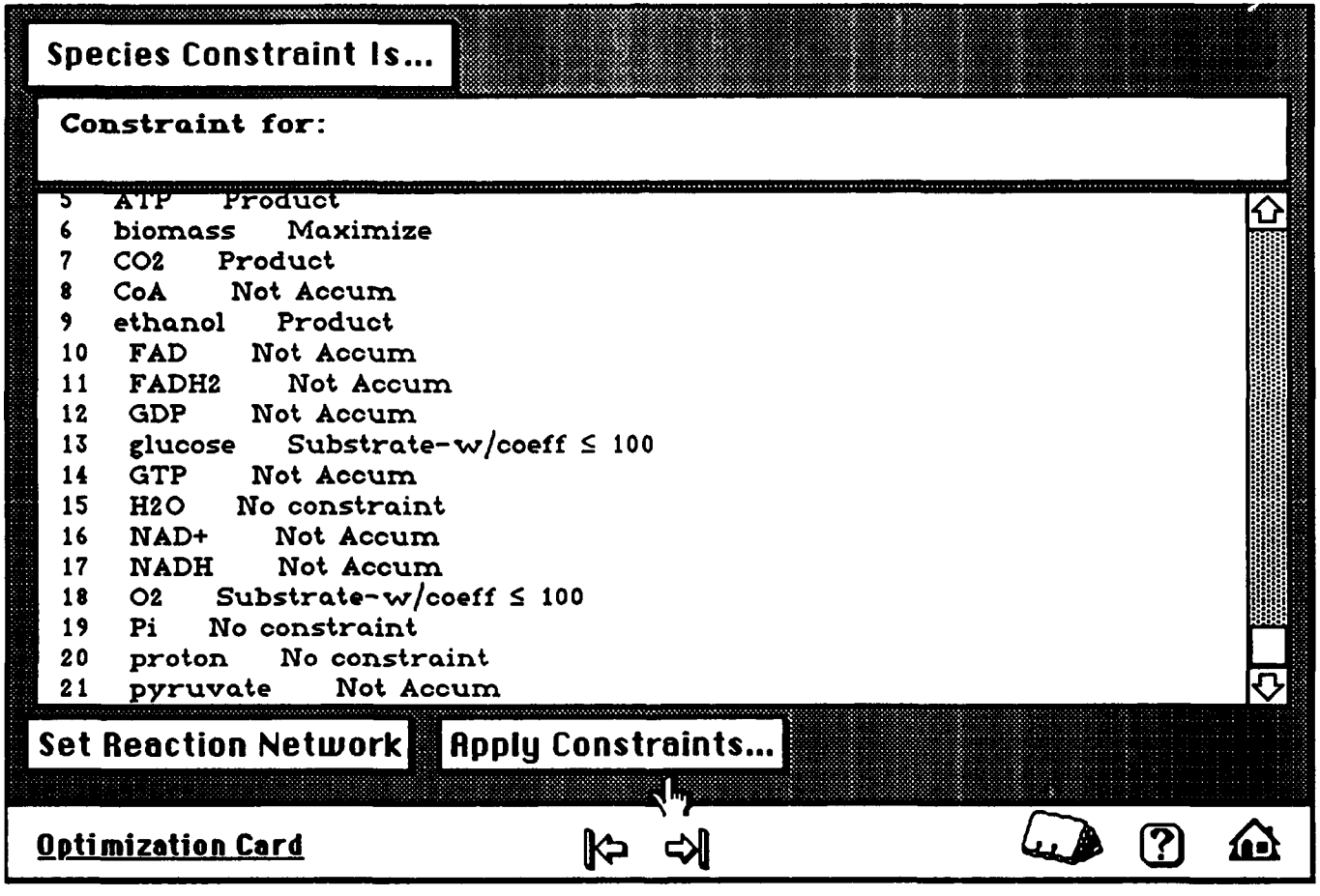


Figure 3.4. Species Constraints

Figure 3.5 depicts the optimization card in a mode for setting constraints on the rates of the reactions in the network. As in setting constraints on the species coefficients, all actions are menu driven. Constraints set here may consist of either equalities or one-sided inequalities.

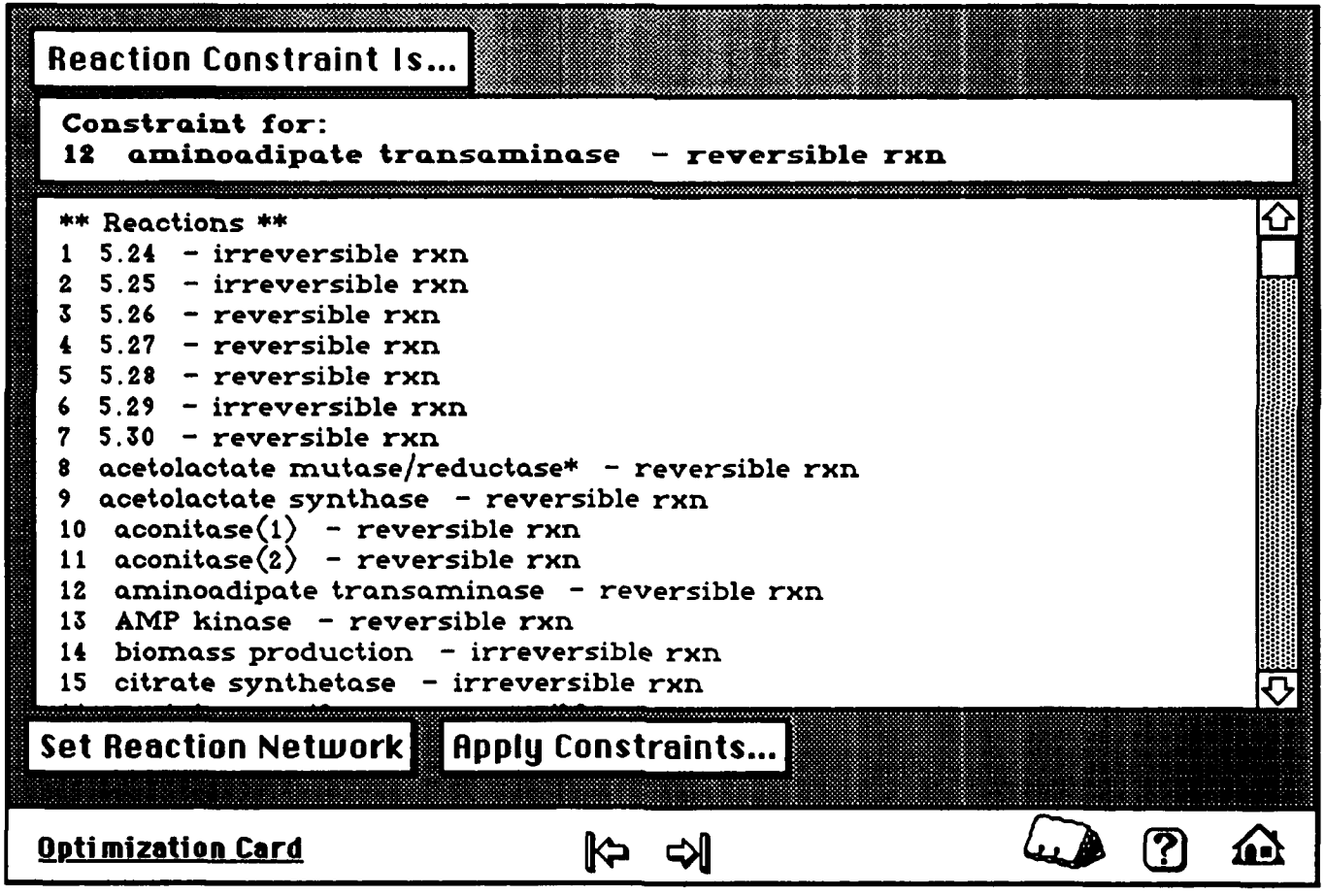


Figure 3.5. Reaction Constraints

Figure 3.6 depicts part of a typical set of results that would be returned from the optimization of a particular system.

The first lines indicate any major diagnostic messages which MINOS may have generated in the course of attempting to solve the linear program representing the reaction system. While this result indicates a good solution, other messages might indicate that the system is infeasible, unbounded, or that the solution is not entirely unique.

The substrates and products listed are the substrates and products with their stoichiometric coefficients in the net reaction of the optimized system.

The results field is usually rather long, with additional data elements consisting of the optimized relative reaction rates as well as the full sets of constraints under which the system was optimized. Thus this field, which can be written off to a text file, provides a relatively complete record of the problem statement as well as its solution.

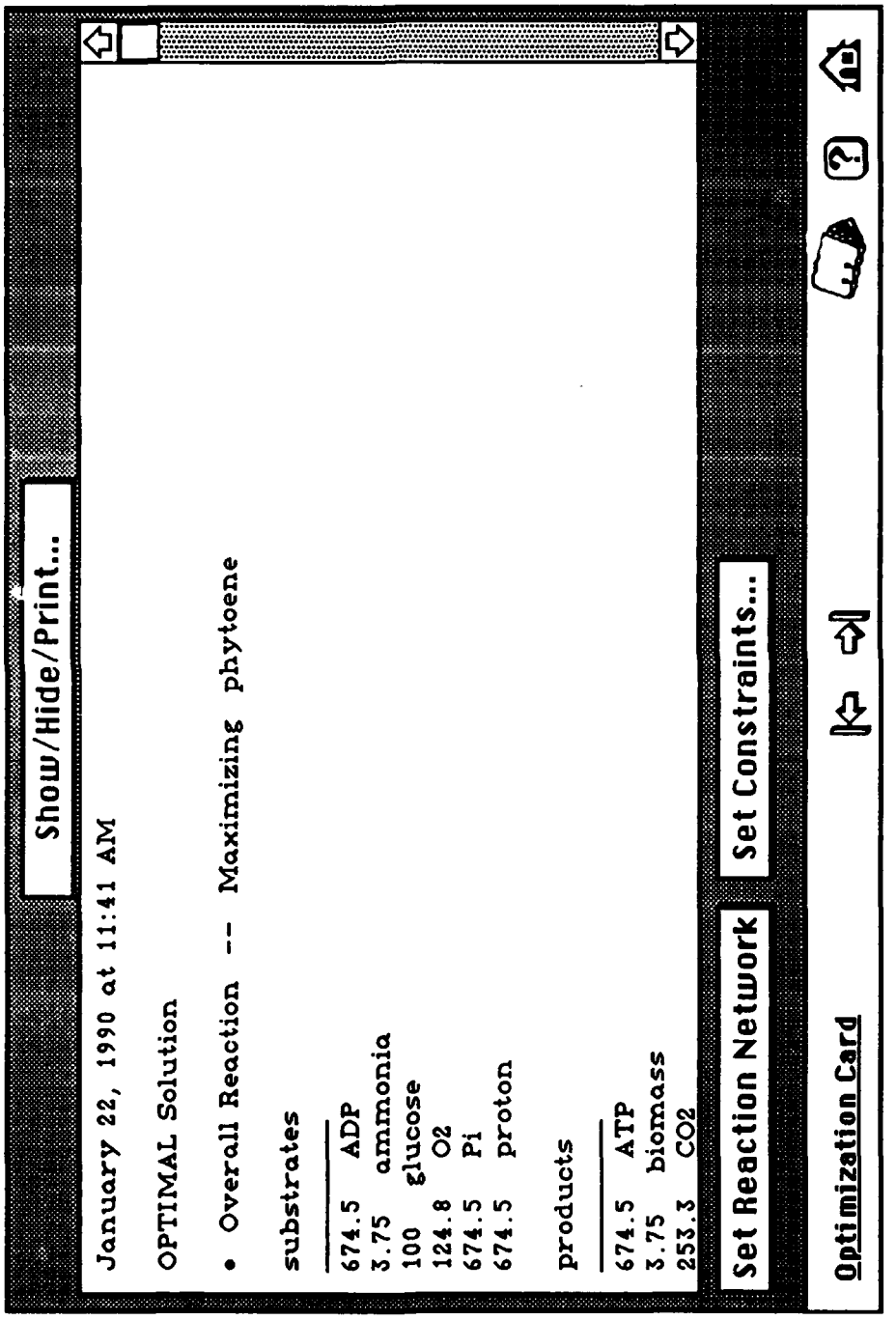
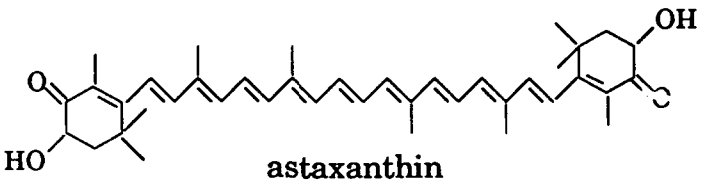


Figure 3.6. Results Example

Chapter 4. Case Study I: Astaxanthin Biosynthesis

4.1. Carotenoid Uses

The production of the carotenoid pigment astaxanthin is of increasing commercial interest. Astaxanthin is a naturally occurring red pigment of molecular formula $C_{40}H_{52}O_4$:



astaxanthin

In nature, C_{40} compounds (carotenoids) are responsible for the red pigmentation of a variety of fish and crustaceans and some other animals, perhaps most notably the pink flamingo (Johnson and Lewis, 1979). Commercial interest in this compound stems from the assumption that most higher animals derive the pigment from food sources. The actual biosynthesis of carotenoids probably occurs in fungi, algae, and small crustaceans (An et al., 1988). Okagbue and Lewis point out some important markets for the carotenoid astaxanthin in particular (1985). In farm raised trout and salmon, diets for the fish must be supplemented with pigments in order to create fish with the pink flesh common to fish caught in the wild, which are more acceptable to consumers than the white fleshed

fish which would result from a diet not supplemented with pigment. Carotenoid supplemented feed for poultry could be useful in giving a more pleasing color to egg yolks and to the chicken flesh. A wide variety of other uses for carotenoid compounds are presented in reviews by Kläui (1981) and Bauernfeind et al. (1971).

4.2. Astaxanthin Production

Astaxanthin could be a particularly important pigment for the fish and poultry industry as it is seems to be a major pigment in salmon and trout (An et al., 1989), and may be useful as a pigment source in poultry products (Johnson et al., 1980). Currently it is produced by one supplier, F. Hoffmann–La Roche and Company (An et al., 1989), by organic synthesis. There is a significant possibility for the lower cost production of the pigment by a biological route. In particular a yeast, *Phaffia rhodozyma*, was discovered in the 1970's (Miller, 1976) which produces carotenoid pigments. Astaxanthin is the major carotenoid present, comprising approximately 85 weight percent of the carotenoids in the cell (Andrewes et al., 1976). In wild type *P. rhodozyma* the concentration of astaxanthin is a rather low 300 to 450 microgram pigment per gram dry cell mass. Researchers are currently working to produce higher intracellular concentrations of the pigment in the yeast (Johnson and Lewis, 1979; An et al., 1989).

An important consideration in the commercial production of astaxanthin using yeast is what the limits of product yield might be. A first guess would be that yield could be greatly improved over the wild-type yields if the organism could be directed to maximize pigment production. The pathway in this example is relatively linear and more readily comprehensible than some more elaborate biochemical systems, thus the calculation of the potential optimum yield is relatively straightforward. The calculation might be done without computer assistance, but it presents a useful, simple example for the types of calculations that can be rapidly done with the Pathway Toolbox Program.

4.3. Biosynthetic Pathway for Astaxanthin

The details of astaxanthin biosynthesis in *Phaffia rhodozyma* have not been specifically elucidated. Available data suggests that the pathway conforms to the general scheme for isoprenoid and carotenoid biosynthesis found in a wide variety of organisms (Andrewes et al., 1975; Britton, 1976, 1983; Johnson and Lewis, 1979; Simpson and Chichester, 1979). For this analysis, the carbon source for growth is assumed to be glucose. Growth on other sugar carbon sources is possible (Miller et al., 1976), with the overall pathway differing most probably in the method of formation of acetyl-CoA.

4.3.1. Isoprenoid Formation

The pathway for the biosynthesis of isopentenyl pyrophosphate, the building block for carotenoid compounds, is diagrammed in Figure 4.1. The diagrammed reactions show the major chemical species. The reactions do not show protons and water, and only one form of the reduction/oxidation compounds, such as NADH, is shown.

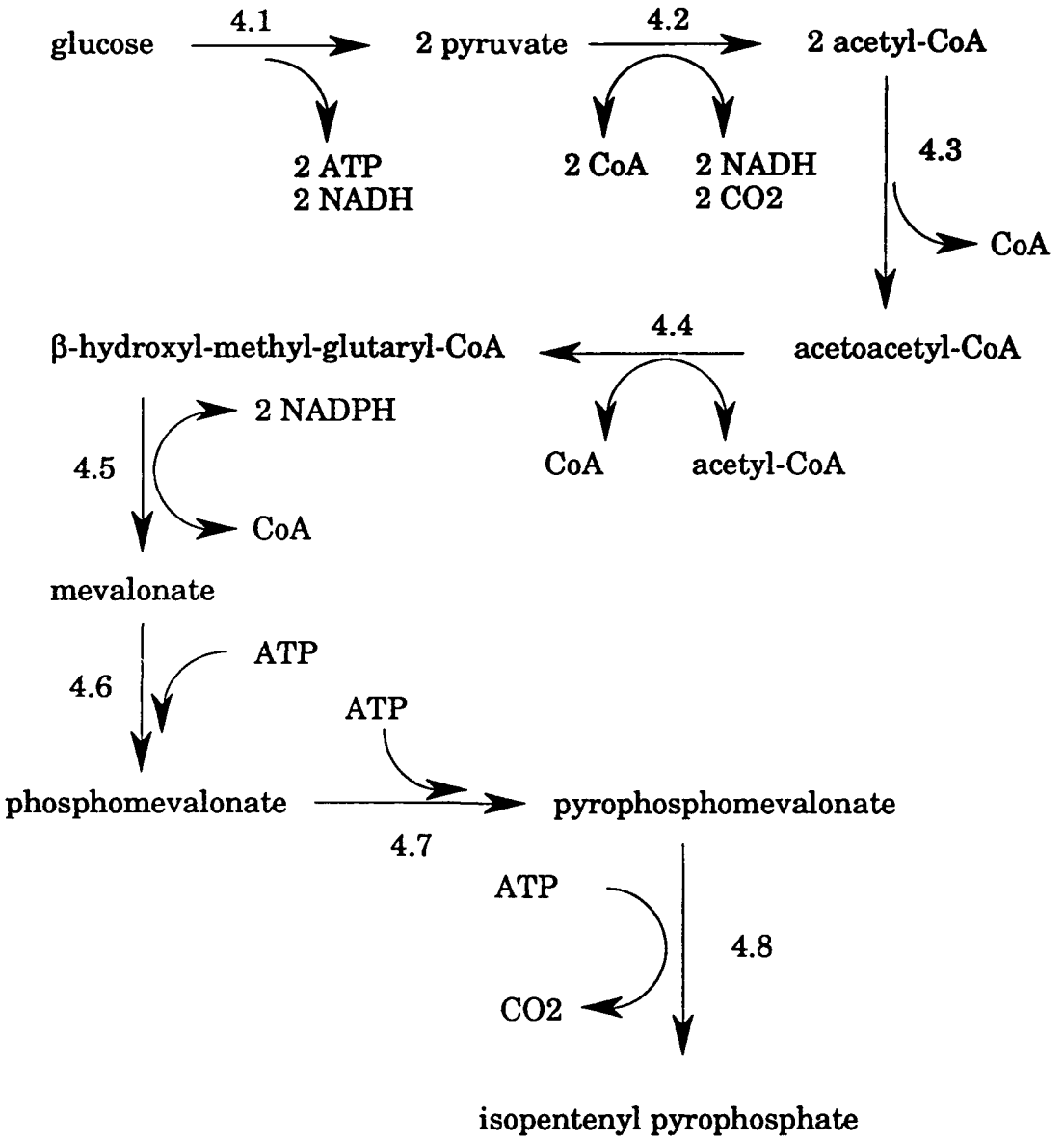


Figure 4.1. Biosynthesis of Isopentenyl Pyrophosphate

Key to reactions in Figure 4.1: 4.1, Embden-Meyerhof pathway; 4.2, pyruvate dehydrogenase complex; 4.3, acetyl-CoA acetyl transferase; 4.4, hydroxymethylglutaryl-CoA synthetase; 4.5, hydroxymethylglutaryl-CoA reductase; 4.6, mevalonate kinase; 4.7, phosphomevalonate kinase; 4.8, pyrophosphomevalonate decarboxylase.

4.3.2. Carotenoid Formation

Eight isoprenoid units are then enzymatically combined to form the basic twenty carbon precursor of all carotenoids, geranylgeranyl pyrophosphate. This pathway is outlined in Figure 4.2.

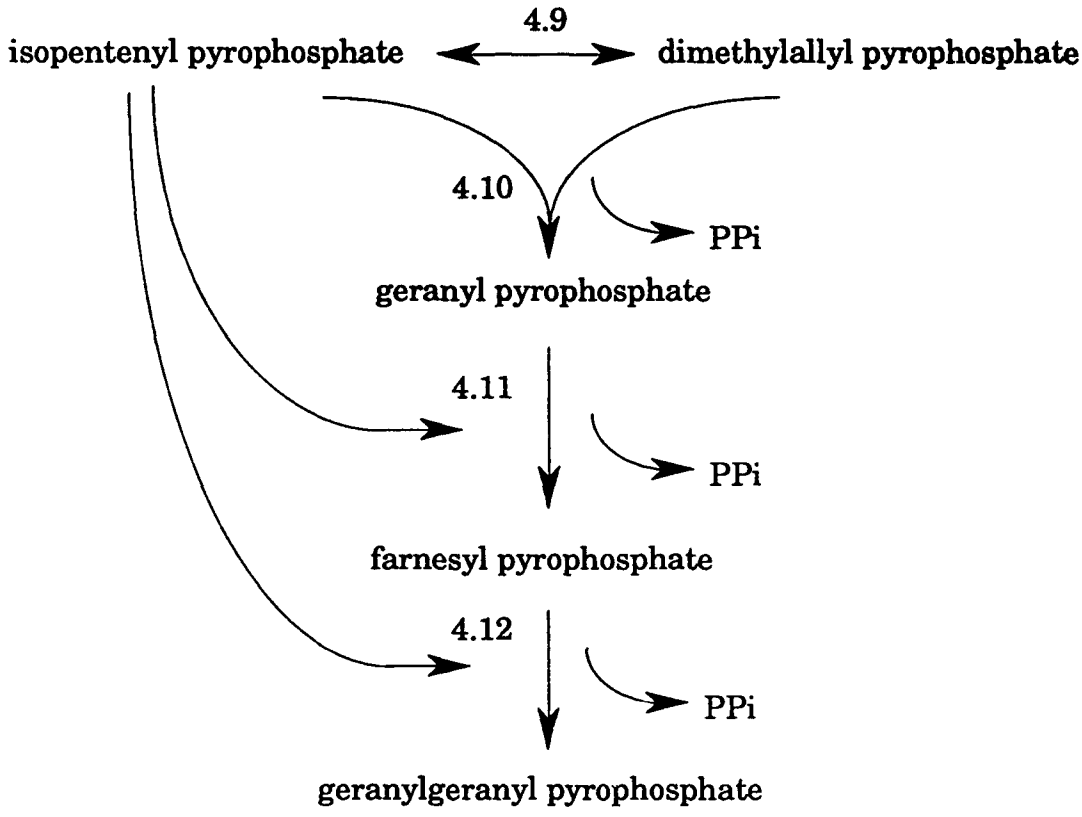


Figure 4.2. Synthesis of Geranylgeranyl Pyrophosphate

Key: 4.9, isopentenyl pyrophosphate isomerase; 4.10 through 4.12, dimethylallyl transferase.

4.3.3. Astaxanthin Formation

Two molecules of geranylgeranyl pyrophosphate combine to form phytoene, and in a series of reduction, isomerization, and oxidation reactions phytoene is transformed to astaxanthin. This pathway is diagrammed in Figure 4.3.

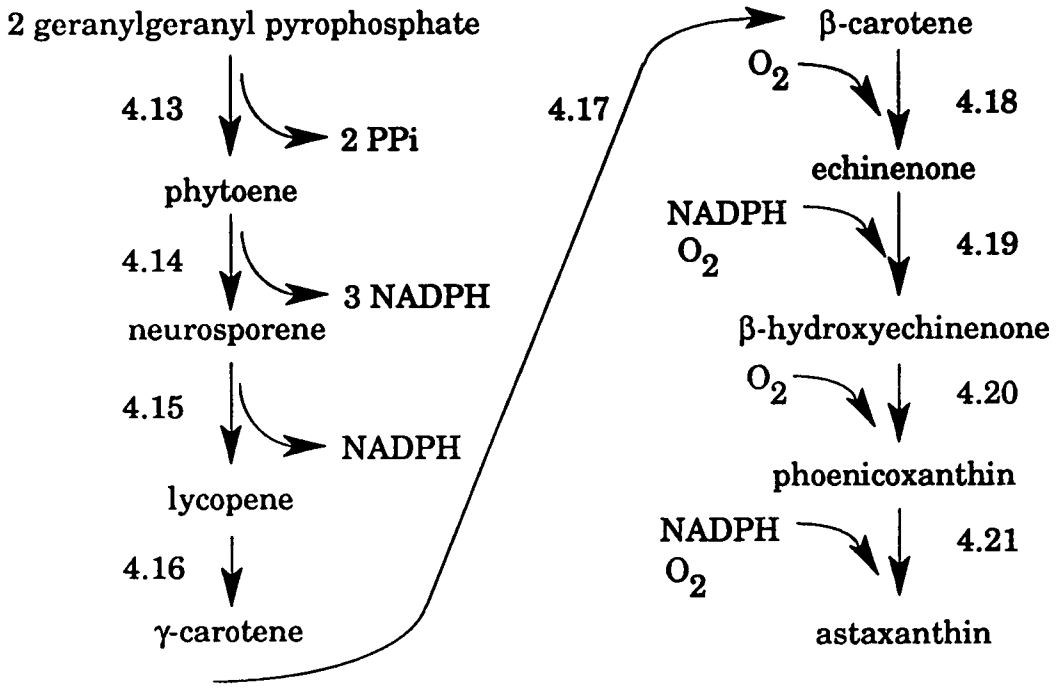


Figure 4.3. Carotenoid Synthesis Reactions

Key (note that these enzyme names serve only as labels to identify the reactions in this analysis and do not necessarily reflect the name of an isolated enzyme): 4.13, phytoene synthetase; 4.14, C_{40} dehydrogenase (1); 4.15, C_{40} dehydrogenase (2); 4.16, C_{40} isomerase (1); 4.17, C_{40} isomerase (2); 4.18, C_{40} oxidase (1); 4.19, C_{40} oxidase (2); 4.20, C_{40} oxidase (3); 4.21, C_{40} oxidase (4).

Figure 4.4 is a diagram of the structures of the 40 carbon chemical species.

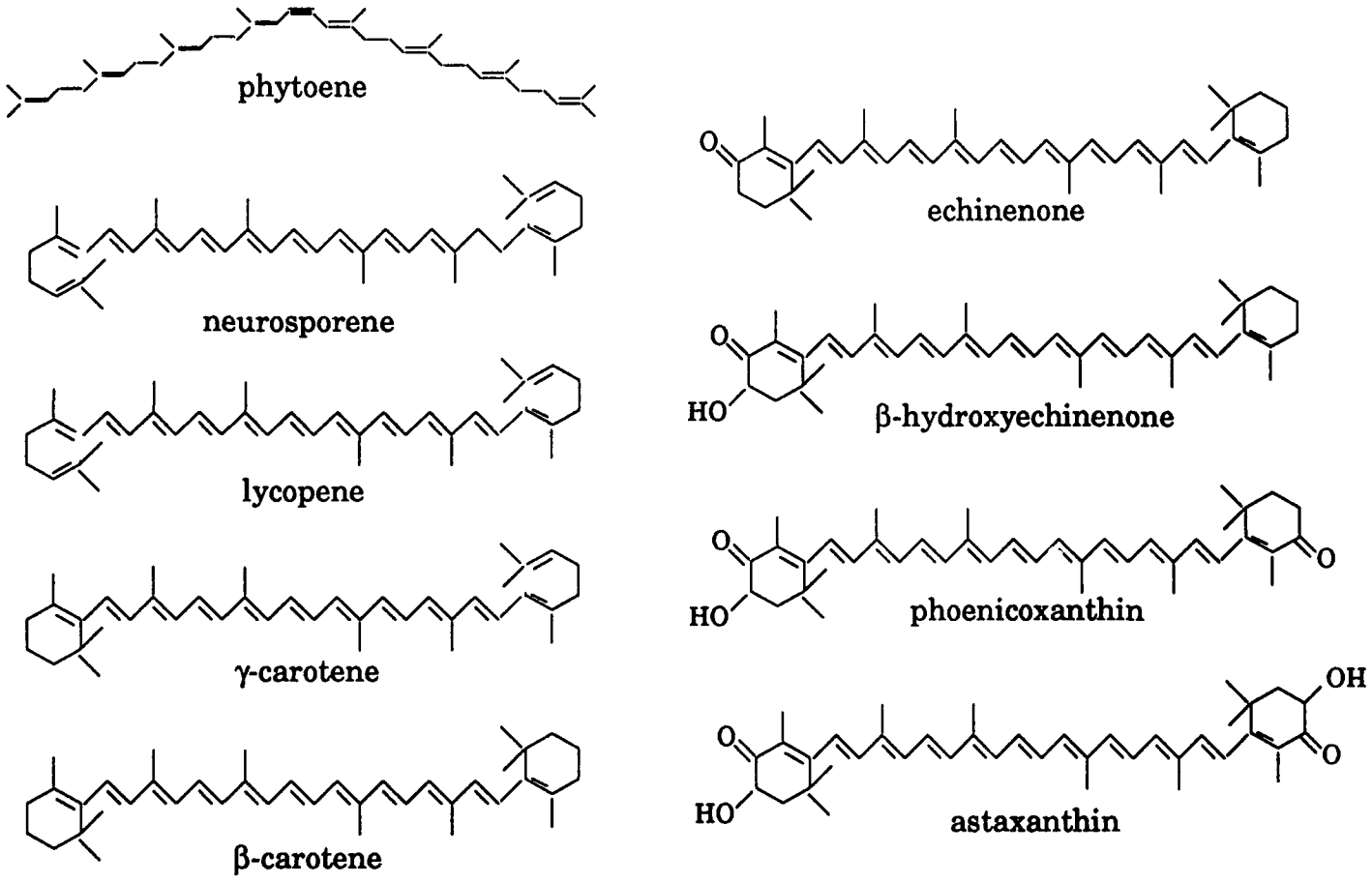


Figure 4.4. Carotenoid Compounds

4.3.4. Other Reactions

To simplify the analysis somewhat, some general reactions are added to the network. These include energy generating reactions, and equilibrium reactions which provide for the interconversion of cellular redox species and of energy carrying nucleotide phosphate species.

The cell can generate energy and reduction power through the tricarboxylic acid (TCA) cycle. For this system, the TCA cycle can be sufficiently represented by a single overall reaction:

$$\text{acetyl-CoA} + 3\,NAD^+ + FAD + GDP + Pi + H_2O = 2\,CO_2 + 3\,NADH + FADH_2 + GTP + CoA + 2\,H^+ \quad (4.22)$$

ATP and GTP may be interconverted through the action of GTP phosphate kinase:

$$ATP + GDP = ADP + GTP \quad (4.23)$$

AMP is phosphorylated by AMP kinase:

$$AMP + ATP = 2\,ADP \quad (4.24)$$

Pyrophosphate, PPi, must be hydrolyzed to organic phosphate, Pi, before it can reenter into the metabolism of the cell in the phosphorylation of ADP to ATP. This hydrolysis is accomplished by a pyrophosphatase:

$$PPi + 2\,H_2O = 2\,Pi + 2\,H^+ \qquad (4.25)$$

If it is assumed that the intracellular redox species may be interconverted by oxidoreductases, the following reactions should be included:

$$FADH_2 + NAD^+ = FAD + NADH + H^+ \qquad (4.26)$$

$$NADP^+ + NADH = NADPH + NAD^+ \qquad (4.27)$$

The activities of these reactions may be set to zero in the quantitative analysis if the effects of these assumptions are to be examined (see also section 4.4.2 and 4.4.3).

Since significant production of astaxanthin occurs only in the presence of oxygen, it may be assumed that the yeast grown aerobically will be able to conduct electron transport phosphorylation. These reactions may be summarized as:

$$FADH_2 + 1/2\ O_2 + 2\ ADP + 2\ Pi + 2\ H^+ = 2\ ATP + FAD + 3\ H_2O \quad (4.28)$$

$$NADH + 1/2\ O_2 + 3\ ADP + 3\ Pi + 4\ H^+ = 3\ ATP + NAD^+ + 4\ H_2O \quad (4.29)$$

Should the conversion of some substrate to biomass need to be examined, an overall reaction representing this conversion is sufficient as long as the conversion of substrate to biomass is a relatively small fraction of the total substrate consumed (Papoutsakis, 1984). A biomass formula, $C_8H_{13}O_4N$, was developed by Minkevich et al. (1978a) which represents a median value of cell elemental compositions and is not greatly varied in most cells. Taking into account regularities in the degree of reductance of cell mass, and an average value of the yield of cell mass on ATP (Stouthamer, 1973; Minkevich et al., 1978a), an equation for the formation of cell mass may be written as:

$$4/3\ \text{glucose} + \text{ammonia} + NADH + 19.8\ ATP + 15.8\ H_2O = \text{biomass} + NAD^+ + 19.8\ ADP + 19.8\ Pi + 18.8\ H^+ \quad (4.30)$$

This reaction does not account for any significant amount of astaxanthin present as an intracellular product.

4.4. Analysis

4.4.1. Carbon Limitation

It would be useful to know what the maximum possible conversion of glucose to the desired product astaxanthin would be in the simplest case of the absence of any limits other than substrate availability. The reactions outlined above involve 44 chemical species. For each of these species and reactions a constraint must be considered.

4.4.1.1. Constraints on Reactions

The constraints for the reactions represent whether or not the reaction is expected to be reversible. The details of the thermodynamics for the reactions are not available, so some biochemistry rules of thumb may be applied. For this network, the only reactions which are assumed to be irreversible are those which involve making or breaking phosphate bonds. This assumption is based on the relatively large changes in energy associated with these bonds. The hydrolysis reactions of ATP to ADP or AMP have free energy changes ($\Delta G^{\circ\prime}$) of -7.60 kcal/mole and -9.96

kcal/mole respectively (Thauer et al., 1977). These constraints are shown in Table 4.1.

Table 4.1. Constraints on Astaxanthin Pathway Reactions

4.1 Overall Embden-Meyerhof pathway	irreversible rxn
4.2 Overall pyruvate dehydrogenase complex	irreversible rxn
4.3 Acetyl-CoA-acetyl transferase	reversible rxn
4.4 Hydroxymethylglutaryl-CoA synthetase	reversible rxn
4.5 Hydroxymethylglutaryl-CoA reductase	reversible rxn
4.6 Mevalonate kinase	reversible rxn
4.7 Phosphomevalonate kinase	reversible rxn
4.8 Pyrophosphomevalonate decarboxylase	reversible rxn
4.9 Isopentenyl pyrophosphate isomerase	reversible rxn
4.10 Dimethylallyltransferase (1)	irreversible rxn
4.11 Dimethylallyltransferase (2)*	irreversible rxn
4.12 Dimethylallyltransferase (3)*	reversible rxn
4.13 Phytoene synthetase*	reversible rxn
4.14 C40 dehydrogenase (1)*	reversible rxn
4.15 C40 dehydrogenase (2)*	reversible rxn
4.16 C40 isomerase (1)*	reversible rxn
4.17 C40 isomerase (2)*	reversible rxn
4.18 C40 oxidase (1)*	reversible rxn
4.19 C40 oxidase (2)*	reversible rxn
4.20 C40 oxidase (3)*	reversible rxn
4.21 C40 oxidase (4)*	reversible rxn
4.22 Overall TCA Cycle	irreversible rxn
4.23 GTP-diphosphate kinase	reversible rxn
4.24 AMP kinase	reversible rxn
4.25 Pyrophosphatase*	irreversible rxn
4.26 NAD/FADH2 transhydrogenase	reversible rxn
4.27 $NAD(P)^+$ transhydrogenase	reversible rxn
4.28 Oxidative Phosphorylation ($FADH_2$)	irreversible rxn
4.29 Oxidative Phosphorylation (NADH)	irreversible rxn
4.30 biomass production	irreversible rxn

4.4.1.2. Constraints on Chemical Species

Most of these species are intracellular intermediates and are thus assumed not to be accumulated or depleted in an overall fermentation reaction. Additionally, the constraints for AMP, GTP, GDP and PPi are set to prevent accumulation of these species, which will leave the energy requirements of the system expressed in terms of ATP and ADP. Since the first effort is to simply determine the carbon limited conversion, no constraints are initially set on the redox species NAD(P)H and $FADH_2$.

It is not possible before hand to determine whether water and protons will be substrates or products, so no constraint is applied to these species.

For this analysis, glucose is assumed to be the only carbon source, with 12 moles available to be used (the coefficient 12 happens to give whole number results). Astaxanthin, the other carotenoid compounds, acetate, biomass, and CO_2 are to be the only possible carbon carrying products, with production of astaxanthin being maximized. For the first analysis, no constraint is applied to the production or consumption of ATP (this constraint consequently also applies to ADP and Pi as the constraints for the three species are not independent). The complete set of constraints is shown in Table 4.2.

Table 4.2. Constraints on Astaxanthin Pathway Chemical Species

Chemical Species	Constraint
1 acetoacetate	Not Accumulated
2 acetoacetyl-CoA	Not Accumulated
3 acetyl-CoA	Not Accumulated
4 ADP	No constraint
5 ammonia	Substrate
6 AMP	Not Accumulated
7 astaxanthin	Maximize
8 ATP	No constraint
9 biomass	Product
10 β-carotene	Product
11 CO_2	Product
12 CoA	Not Accumulated
13 3,3-dimethylallyl.pyrophosphate	Not Accumulated
14 echinenone	Product
15 FAD	No constraint
16 $FADH_2$	No constraint
17 farnesyl.pyrophosphate	Not Accumulated
18 γ-carotene	Not Accumulated
19 GDP	Not Accumulated
20 geranyl.pyrophosphate	Not Accumulated
21 geranylgeranyl.pyrophosphate	Not Accumulated
22 glucose	Substrate-w/coeff ≤ 12
23 GTP	Not Accumulated
24 H_2O	No constraint
25 3-hydroxy-3-methyl-glutaryl-CoA	Not Accumulated
26 3-hydroxyechinenone	Not Accumulated
27 3-isopentenyl.pyrophosphate	Not Accumulated
28 lycopene	Product

29	mevalonate	Not Accumulated
30	NAD^+	No constraint
31	NADH	No constraint
32	$NADP^+$	No constraint
33	NADPH	No constraint
34	neurosporene	Product
35	O_2	Substrate
36	phoenicoxanthin	Product
37	5-phosphomevalonate	Not Accumulated
38	phytoene	Product
39	Pi	No constraint
40	PPi	Not Accumulated
41	proton	No constraint
42	5-pyrophosphomevalonate	Not Accumulated
43	pyruvate	Not Accumulated

4.4.1.3. Results

If the system were to be solved without regard to its internal structure, considering only macroscopic balances, a carbon limited maximum conversion yield would be:

$$6\ 2/3\ \text{glucose} + 22\ H_2 = 1\ \text{astaxanthin} + 36\ H_2O \qquad (4.31)$$

When the constraints described above are applied to the system of reactions with due consideration for it structure, the resulting net reaction for the maximum possible production of astaxanthin from glucose is:

$$12\ \text{glucose} + 4\ O_2 + 14\ \text{NADPH} + 48\ \text{NAD}^+ = 1\ \text{astaxanthin} + 32\ CO_2 + 14\ \text{NADP}^+ + 48\ \text{NADH} + 12\ H_2O + 34\ H^+ \qquad (4.32)$$

This overall reaction is the result of the linear combination of the constituent reactions with steady state rates as shown in Table 4.3. The rate values shown are relative rates. The rates do not represent absolute values but only rates relative to the other rates displayed. For instance, if the reaction catalyzed by C40 dehydrogenase (1)* proceeds at some specific rate, then the reaction catalyzed by Acetyl-CoA-acetyl transferase must proceed at 8 times that specific rate in the optimized pathway.

Table 4.3. Optimized Reaction Rates for Astaxanthin Production

Rate	Reaction
8	Acetyl-CoA-acetyl transferase
1	C40 dehydrogenase (1)*
1	C40 dehydrogenase (2)*
1	C40 isomerase (1)*
1	C40 isomerase (2)*
1	C40 oxidase (1)*
1	C40 oxidase (2)*
1	C40 oxidase (3)*
1	C40 oxidase (4)*
2	Dimethylallyltransferase (1)
2	Dimethylallyltransferase (2)
2	Dimethylallyltransferase (3)*
8	Hydroxymethylglutaryl-CoA reductase
8	Hydroxymethylglutaryl-CoA synthetase
2	Isopentenyl pyrophosphate isomerase
8	Mevalonate kinase
12	Overall Embden-Meyerhof pathway
24	Overall pyruvate dehydrogenase complex
8	Phosphomevalonate kinase
1	Phytoene synthetase*
8	Pyrophosphatase*
8	Pyrophosphomevalonate decarboxylase

The numbers derived from this analysis should be used with some caution as they are sensitive to some degree on the types and numerical values of the constraints used. A sensitivity analysis to analyze the effects of changes in the numerical values of the constraints has not been performed as it is not included in the capability of the MINOS linear programming package used. In the future, this capability could be added by modifying the software to utilize a numerical package which easily performs sensitivity analyses on linear optimization problems, such a version of LINDO. MINOS was selected for use in this project, because it is designed for batch operation while LINDO would have required a greater programming effort to make it effectively work on these problems in batch mode.

The particular stoichiometric coefficients of species other than those which have been specified or are being optimized should not be considered unique. Indeed, one can see by inspection that the interconversion of the various redox species could give rise to an large number of net reactions which produce the same yield of astaxanthin as equation 4.32 yet have widely differing stoichiometric coefficients for the redox species.

In general, a moderately detailed model of a biochemical reaction system will almost inevitably contain reaction cycles which have little or no direct effect on yields of carbon products, yet can greatly influence the energy yields of the system (as measured by ATP yield) and may be able to

interconvert the various chemical species which serve as reduction/oxidation intermediates. Because the linear optimization solution method tries to maximize the objective function, here a single chemical species, through the reaction rates, a solution may be reached which yields an optimum values of that function yet is not unique in terms of the reaction rates. That is, other rates may give rise to the same optimal value of the desired species. The Pathway Toolbox Program warns the user of this occurrence when the MINOS output file indicates its presence; however, MINOS (and from experience in this project, LINDO), does not always detect this problem. Consequently, the user has to exercise some judgement and caution, derived from familiarity with the biochemical system, in interpreting the results of the optimization process.

It is clear that in the synthesis of astaxanthin from glucose in this pathway, the maximum carbon limited yield of product is one mole of astaxanthin per twelve moles of glucose. There is also a large amount of reducing power generated in the reaction with 34 more reducing species molecules being produced than consumed. Since cells are not expected to accumulate these reducing species, a further analysis should be conducted.

4.4.2. Reduction/Oxidation Limitations

4.4.2.1. Constraints on Reactions

For the analysis of the system with constraints applied to the redox species, no constraints on the directions or magnitudes of the reaction rates beyond those of the previous analysis are needed.

4.4.2.2. Constraints on Chemical Species

The same constraints for the species are used as in the previous analysis with the additional constraint that the intracellular reduction and oxidation species such as NADH, NADPH, and $FADH_2$ are not allowed to accumulate.

4.4.2.3. Results

A solution to the astaxanthin synthesis system with the added constraints on the redox species is:

$$12\ \text{glucose} + 21\ O_2 + 102\ \text{ADP} + 102\ \text{Pi} + 102\ \text{proton} = 1\ \text{astaxanthin} + 32\ CO_2 + 102\ \text{ATP} + 148\ H_2O \tag{4.33}$$

However, an equally valid solution is given by the net reaction:

$$12\ \text{glucose} + 21\ O_2 + 68\ \text{ADP} + 68\ \text{Pi} + 68\ \text{proton} = 1\ \text{astaxanthin} + 32\ CO_2 + 68\ \text{ATP} + 114\ H_2O \quad (4.34)$$

The yield of product on glucose in these two cases is the same as that derived from carbon limitations alone. The yield of ATP differs in equations 4.33 and 4.34. These two results represent the extremes of two different cases involving the fate of the redox species. There is only a small difference in the reaction network underlying equation 4.33 and that behind 4.34 which arises, as discussed above, from the problem of non-unique optima.

The yield of 68 ATP is derived from first using the NADH to reduce FAD to $FADH_2$ which is then used for electron level phosphorylation of ADP with a yield of two ATP per $FADH_2$. This was the first solution the numerical routine arrived at. It may be somewhat unrealistic in the high activity of the $FADH_2$:NADH oxidoreductase which drives all the reducing power through the electron level phosphorylation process in the form of $FADH_2$.

To arrive at the yield of 102 ATP in equation 4.33, the NADH generated in the reaction system is all run through oxidative

phosphorylation yielding three ATP per NADH. There are several methods by which this answer may be derived. Using the information from the original optimization run for unconstrained carbon product yield (equation 4.32), the coefficients of the carbon carrying species could be specified at the optimized levels, and then a run could be made optimizing ATP production. That method was used to arrive at equation 4.33. A second method would be simply to set the rate of the $FADH_2$:NADH oxidoreductase to zero, corresponding to a biological condition of no activity for that enzyme in the system.

ATP yields between these two values could be obtained in the cell by running varying ratios of $FADH_2$ and NADH going through the electron level phosphorylation process. Of course, ATP yields below both of the values could be obtained through the addition of some futile cycle to the reaction system. These futile cycles would serve only to hydrolyze ATP without linking the energy released to another reaction.

In terms of energy yields, what equations 4.33 and 4.34 indicate is that the production of astaxanthin can be a energy producing process for the cell. The optimal yield of energy is below that which might be achieved by utilizing the substrate for energy production alone (a cell might aerobically generate 36 ATP per glucose in contrast to the 8.5 yield of equation 4.33). Since the process yields energy for the cell, it should be theoretically possible to very closely approach the maximum calculated

astaxanthin yield. If maintenance and growth energy requirements could be kept low enough to be met with the energy generated by the process, no substrate would be required to be diverted into energy production pathways. Under these conditions, cells with a low growth rate, with corresponding low carbon demand for cell mass, could be high yield producers of astaxanthin.

In any case, a second important aspect of the reaction system with the reduction constraints in both examples is that it requires a much greater oxygen to astaxanthin ratio, 21:1, than the system with no reduction constraints which has a 4:1 ratio. This additional oxygen is required as an electron sink to allow the cell to dump off the electrons created in the oxidation of the substrate, glucose, to the astaxanthin product. This high oxygen demand foreshadows possible problems with oxygen transfer in fermentation. The cells must be able to uptake 1.75 moles of oxygen with every mole of glucose to maintain their redox balance. This is somewhat greater than even the requirement for 1.53 moles oxygen and one mole glucose for each mole of biomass formed in aerobic growth. The numbers also suggest a possible avenue around this problem. For instance, if a pathway to an electron acceptor other than oxygen were available for the oxidation of the reduced species generated by the astaxanthin pathway, then the cell would only require 1/3 molecule of oxygen per glucose molecule consumed, greatly reducing the oxygen transfer problem.

4.4.3. Other Constraints

Other types of solutions for the synthesis network are available, one of which is to use part of the glucose substrate rather than oxygen as an electron sink. For instance, an additional constraint might be added that the coefficient of oxygen in the overall reaction is not greater than 4, the value in equation 4.32. This constraint reflects some level of oxygen limitation, a situation which is often present in production scale bioreactors. The net reaction maximizing astaxanthin production in this case would be:

$$\begin{aligned} 6.001\ \text{ammonia} + 108.2\ \text{ATP} + 12\ \text{glucose} + 74.82\ H_2O + 4\ O_2 = \\ 108.2\ \text{ADP} + 0.3334\ \text{astaxanthin} + 6.001\ \text{biomass} + \\ 10.67\ CO_2 + 108.2\ \text{Pi} + 108.2\ \text{proton} \end{aligned} \qquad (4.35)$$

The production of biomass in this equation is a reflection of the fact that in this reaction network, biomass is the chemical species which can be created from glucose and used as an electron sink. While biomass was allowed as a possible product in all the earlier cases, its production was not needed to get a maximal yield in those previous cases. If a more elaborate model were to be constructed which included other metabolic products, such as acetate which is a common microbial product, one of those products might serve as an electron sink.

One difficulty with introducing biomass production into the system is that it will complicate the understanding of the linear optimization process. An enzyme generates products at some rate, but the total amount of enzyme present in a system is proportional to the total cell mass present (assuming a constant concentration of enzyme per mass of cell). Thus the total rate of production of a product, P, in a system by an enzyme with specific rate rate r_p is:

$$\frac{dP}{dt} = r_p X \qquad (4.35)$$

where X is the total cell mass in the system. If biomass is being produced, then X is also a function of time, and the resulting problem description is no longer linear. However, the fundamental assumption in the linear analysis of the system is that it is an open system. The feed stream (substrates) flows into the system, and the products stream flows out of it. Thus the biomass produced does not alter the linear behavior of the model. Of course, the results of this type of model are insufficient to describe a batch type of reaction, but they do allow a clearer understanding of the nature of the process within the systems' constituent cells.

Equation 4.35 indicates that there is a net consumption of ATP in the reaction. Since in growth on glucose, the cell uses glucose to produce the ATP, an additional constraint should be considered to get a more

correct picture. ATP should not be consumed in the net reaction of the system. This additional constraint generates a net reaction of:

$$3.83\ \text{glucose} + 0.9537\ \text{ammonia} + 4\ O_2 = 0.2132\ \text{astaxanthin} + 0.9537\ \text{biomass} + 6.822\ CO_2 + 12.67\ H_2O \qquad (4.36)$$

In equation 4.36, the solution which maximizes the amount of astaxanthin produced no longer involves consumption of the entire quantity of substrate available, rather the net reaction is oxygen limited. What can be noted from this result is that even in this oxygen limited reaction, the ratio of astaxanthin to biomass production is still much higher, approximately 2:9, than what is seen in wild type growth where the ratio is approximately 1:3300.

4.5. Summary

It is now apparent that in even this relatively straightforward linear reaction network, the maximum yield of the desired product can change significantly once a small number of constraints are added to the system. Even the consideration of only the structure of the pathway itself decreases the maximum yield of astaxanthin on glucose 1 mole astaxanthin per 6 2/3 moles glucose to 1 mole astaxanthin per 12 moles glucose. However, it is also fairly clear that if the biochemical reaction system model presented

here is correct, then the potential yield of astaxanthin on glucose could be dramatically increased in comparison to current yields.

The nature of the specific changes in growth environment and organism metabolism are beyond the scope of this effort. The optimized reaction rates derived here (Table 4.3) do provide some indication of the direction in which modifications need to proceed, as the rates provide information on what relative levels of enzyme activities are needed in order to maximize production of astaxanthin.

Chapter 5. Case Study II: Penicillin Biosynthesis

5.1. Overview of the Process

The production of the antibiotic penicillin has been an important fermentation process for more than forty years. The yield of penicillin from feed stock has been extensively studied, and a paper by Cooney and Acevedo (1977) summarizes the general form of the biosynthetic pathway for penicillin G (benzylpenicillin) in *Penicillium chrysogenum* . An analysis of this pathway using linear optimization techniques provides a good exercise in checking the results obtained by those techniques with the results obtained by Cooney and Acevedo. Because the pathway is more complex than that for the synthesis of astaxanthin, it also demonstrates the utility of the optimization approach in understanding a pathway which is hard to quantitatively analyze by hand.

5.2. Pathway for Penicillin Biosynthesis

The pathway analyzed here is that of Cooney and Acevedo (1977) as adapted from Demain (1966). To simplify the description of the pathway, it is broken into five segments. Four of these synthesize precursor components and the fifth combines them to produce the final penicillin product. The pathways are constructed assuming glucose is the primary carbon and energy source for the cells, sulfate is the sulfur source,

ammonia is the nitrogen source, and phenylacetic acid is provided for as a biosynthetic precursor.

5.2.1. Cysteine Biosynthesis

The pathway for the production of cysteine is shown in Figure 5.1. Certain reactions shown are actually the net of several individual reactions for a non-branched series of reactions. These were lumped in the original references, and the names given to them here are notional and used only for identification in the optimization routine. These notional names are marked by an asterisk.

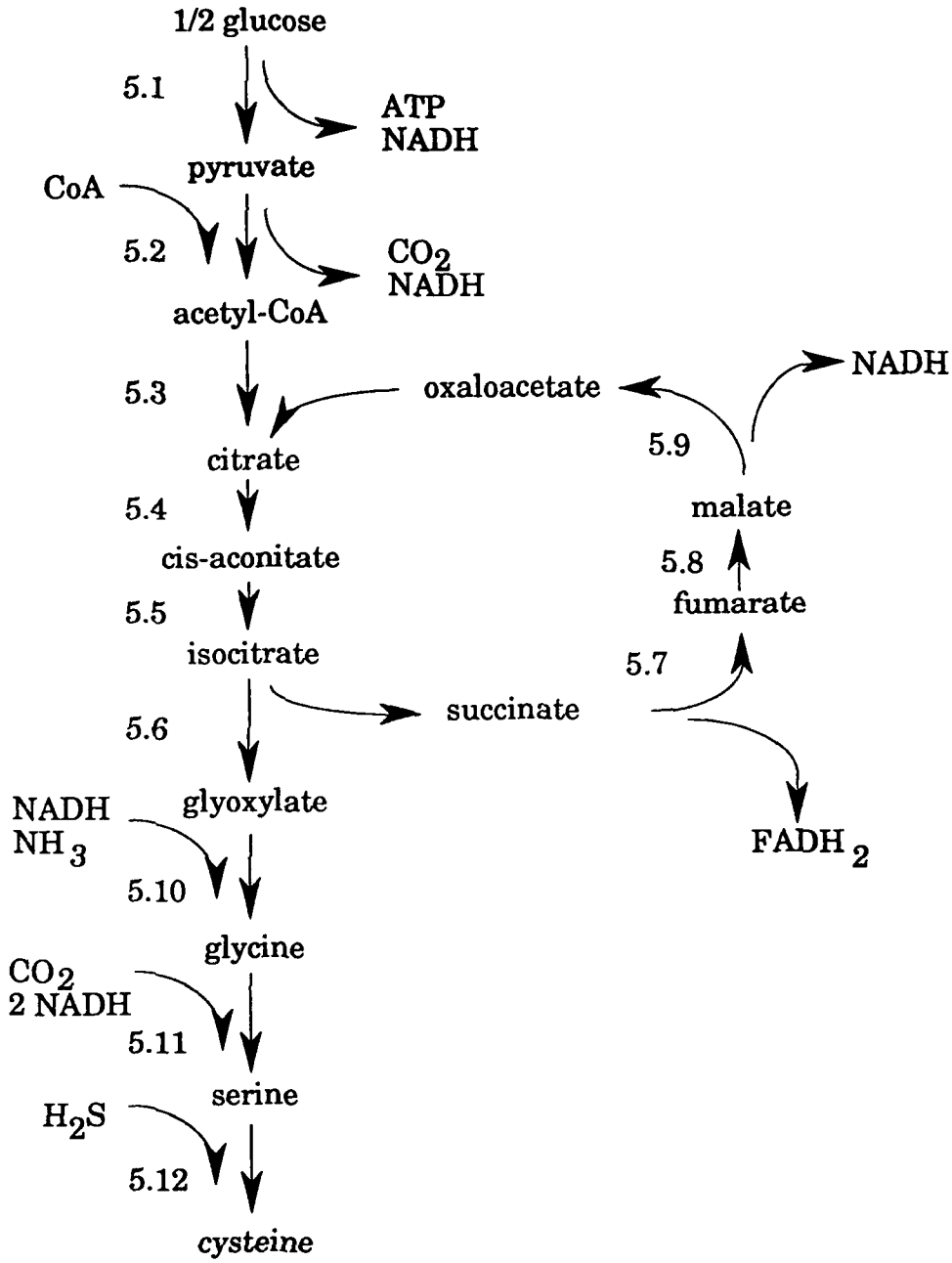


Figure 5.1. Cysteine Pathway

(adapted from Cooney and Acevedo, 1977)

Key to reactions in Figure 5.1: 5.1, Embden-Meyerhof pathway; 5.2, pyruvate dehydrogenase complex; 5.3, citrate synthase; 5.4 and 5.5, aconitase; 5.6, isocitrate lyase; 5.7, succinate dehydrogenase; 5.8, fumarase; 5.9, malate dehydrogenase; 5.10, transaminase*; 5.11, serine hydroxymethyltransferase; 5.12, cysteine synthase

Figure 5.2 illustrates the biological reduction of sulfate to sulfide for cysteine synthesis. This pathway is included in the analysis as a net reaction:

$$\text{sulfate} + 2\,\text{ATP} + 4\,\text{NADH} + 4\,H^{+} = \text{sulfide} + \text{ADP} + \text{AMP} + 4\,NAD^{+} + 3\,\text{Pi} + H_2O \qquad (5.13)$$

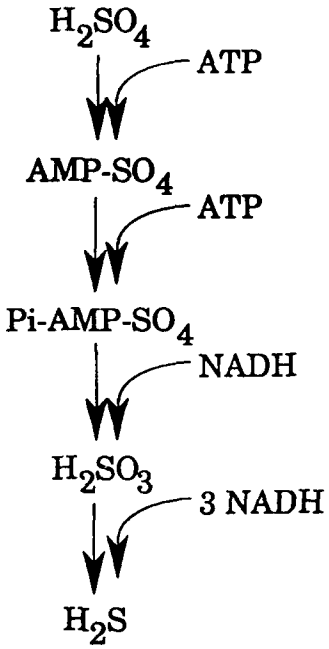


Figure 5.2. Sulfate Reduction

(adapted from Cooney and Acevedo, 1977)

5.2.2. Valine Biosynthesis

Figure 5.3 outlines the pathway for the biosynthesis of the amino acid valine. An aspect of this pathway to note is that reaction 5.15 includes the cofactor NADH. The original analysis of the pathway by Cooney and Acevedo (1977) does not include this source of reducing power, however, a check of the carbons and available electrons in the reaction indicates that two electrons are required to balance the reaction. NADH is used here as a generic source of reducing power, providing two electrons per molecule of NADH.

Again notional names are indicated with an asterisk in the reaction key.

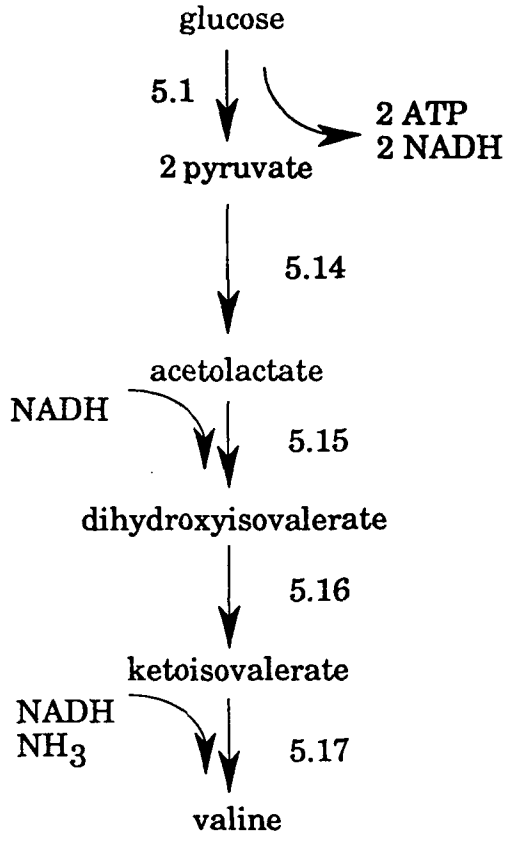


Figure 5.3. Valine Pathway

(adapted from Demain, 1966 and Lehninger, 1975)

Key: 5.1, Embden-Meyerhof pathway; 5.2, pyruvate dehydrogenase complex; 5.14, acetolactate synthase;

5.15, acetolactate mutase/reductase*; 5.16, dihydroxyacid dehydratase;

5.17, valine transaminase

5.2.3. α-aminoadipate Synthesis

Figure 5.4 outlines the pathway for the synthesis of α-aminoadipate from a glucose feedstock. Again notional names are indicated with an asterisk in the reaction key.

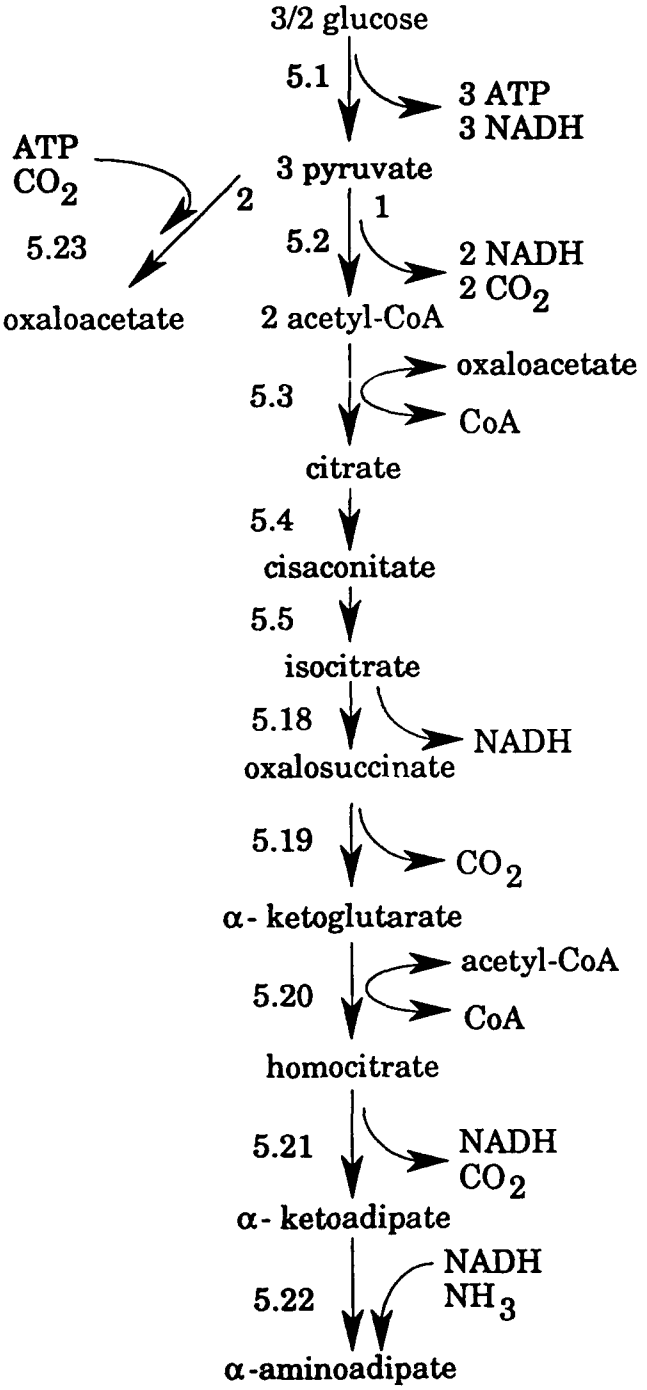


Figure 5.4. α-aminoadipate Synthesis
(adapted from Cooney and Acevedo, 1977)

Key to reactions in Figure 5.4: 5.1, Embden-Meyerhof pathway; 5.2, pyruvate dehydrogenase complex; 5.3, citrate synthase; 5.4 and 5.5, aconitase; 5.18-5.19, α-ketoglutarate dehydrogenase; 5.20, homocitrate synthase; 5.21, homocitrate decarboxylase*; 5.22, aminoadipate transaminase; 5.23, pyruvate carboxylase

5.2.4. Penicillin Assembly

Figure 5.5 outlines the pathway which synthesizes penicillin from the constituent substrates which have been synthesized from glucose or provided in the feedstock. Rather than create names for these reactions, they are manipulated using the reaction number shown as an identifier.

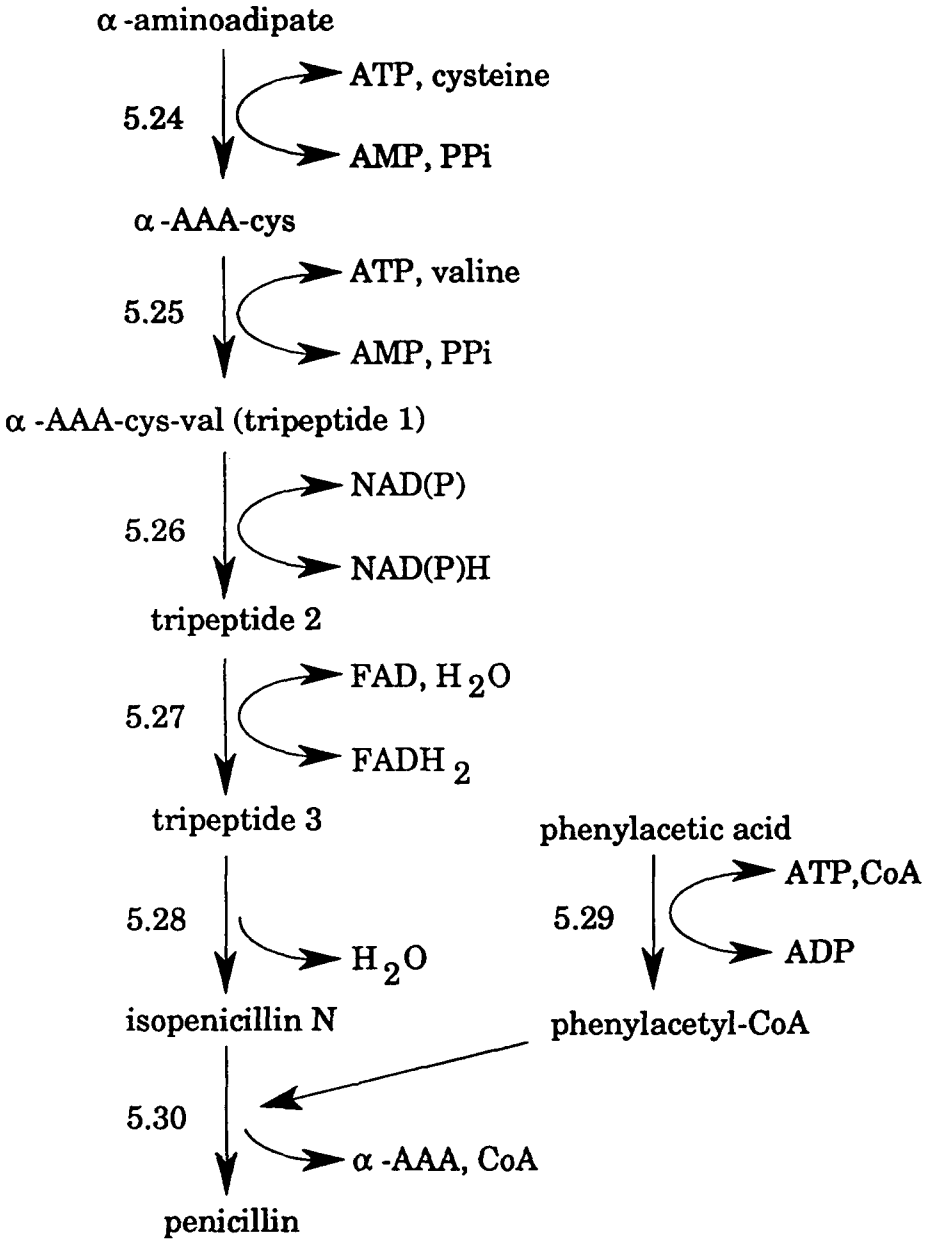


Figure 5.4. α-aminoadipate Synthesis
(adapted from Cooney and Acevedo, 1977)

5.2.5. Other Reactions

As done in the first case study, some general reactions are added to the network to simplify the analysis to some degree. These include energy generating reactions, and equilibrium reactions which provide for the interconversion of cellular redox species and of energy carrying nucleotide phosphate species.

The cell can generate energy and reduction power through the tricarboxylic acid (TCA) cycle. For this system, the TCA cycle can be sufficiently represented by a single overall reaction:

$$\text{acetyl-CoA} + 3\ NAD^+ + FAD + GDP + Pi + H_2O = 2\ CO_2 + 3\ NADH + FADH_2 + GTP + CoA + 2\ H_+ \quad (5.31)$$

ATP and GTP may be interconverted through the action of GTP phosphate kinase:

$$ATP + GDP = ADP + GTP \quad (5.32)$$

AMP is phosphorylated by AMP kinase:

$$AMP + ATP = 2\ ADP \quad (5.33)$$

Pyrophosphate, PPi, must be hydrolyzed to organic phosphate, Pi, before it can reenter into the metabolism of the cell in the phosphorylation of ADP to ATP. This hydrolysis is accomplished by a pyrophosphatase:

$$PPi + 2\,H_2O = 2\,Pi + 2\,H^+ \qquad (5.34)$$

If it is assumed that some of the intracellular redox species may be interconverted by oxidoreductases, the following reactions should be included:

$$NADP^+ + NADH = NADPH + NAD^+ \qquad (5.35)$$

The activities of this reaction may be set to zero in the quantitative analysis if the effects of this assumption are to be examined.

In their analysis, Cooney and Acevedo imply the presence of an operating tricarboxylic acid (TCA) cycle. To reduce the number of reactions to be modeled, the net reaction of a single pass through the TCA cycle may be summarized in a single reaction:

$$\text{acetyl-CoA} + GDP + Pi + 3\,NAD^+ + FAD + 2\,H_2O = 2\,CO_2 + GTP + 3\,NADH + FADH_2 + CoA + 3\,H^+ \qquad (5.36)$$

Since significant production of penicillin occurs aerobically, it may be assumed that electron transport phosphorylation takes place. Following the assumption of Cooney and Acevedo (1977) that the oxidation of one mole of NADH yields two moles of ATP, this reaction may be summarized as:

$$NADH + 1/2\ O_2 + 2\ ADP + 2\ Pi + 3\ H^+ =$$
$$2\ ATP + NAD^+ + 3\ H_2O \qquad (5.37)$$

Cooney and Acevedo find it unlikely that $FADH_2$ is utilized for energy production in the fungi producing penicillin, and assume instead that it is oxidized to yield hydrogen peroxide which is then oxidized to form water and oxygen with the net reaction being:

$$FADH_2 + O_2 = 2\ H_2O \qquad (5.38)$$

Should the conversion of some substrate to biomass need to be examined, reaction 4.30 from the first case study also can be used in this case:

$$4/3\ glucose + ammonia + NADH + 19.8\ ATP + 15.8\ H_2O =$$
$$biomass + NAD^+ + 19.8\ ADP + 19.8\ Pi + 18.8\ H^+ \qquad (5.39)$$

5.3. Yield Analysis

5.3.1. Carbon Limitations

5.3.1.1. Constraints on Reactions

The constraints on the reversibility or irreversibility of the reactions in the network are taken from Cooney and Acevedo (1977). They are enumerated in Table 5.1.

Table 5.1. Reversibilities of Reactions

1 5.24	irreversible rxn
2 5.25	irreversible rxn
3 5.26	reversible rxn
4 5.27	reversible rxn
5 5.28	reversible rxn
6 5.29	irreversible rxn
7 5.30	reversible rxn
8 acetolactate mutase/reductase*	reversible rxn
9 acetolactate synthase	reversible rxn
10 aconitase(1)	reversible rxn
11 aconitase(2)	reversible rxn
12 aminoadipate transaminase	reversible rxn
13 AMP kinase	reversible rxn
14 biomass production	irreversible rxn
15 citrate synthetase	irreversible rxn
16 cysteine synthase	reversible rxn
17 dihydroxyacid dehydratase	reversible rxn
18 fumarase	reversible rxn

19	GTP-diphosphate kinase	reversible rxn
20	homocitrate decarboxylase*	reversible rxn
21	homocitrate synthase	reversible rxn
22	hydrogenase	reversible rxn
23	α-ketoglutarate dehydrogenase	reversible rxn
24	isocitrate lyase	irreversible rxn
25	malate dehydrogenase	reversible rxn
26	$NAD(P)^+$ transhydrogenase	reversible rxn
27	NADH ferredoxin oxidoreductase	reversible rxn
28	Overall Embden-Meyerhof pathway	irreversible rxn
29	Overall pyruvate dehydrogenase complex	irreversible rxn
30	Overall TCA Cycle	irreversible rxn
31	oxidation of $FADH_2$ (net)	irreversible rxn
32	Oxidative Phosphorylation (mit/mal-asp NADH)	irreversible rxn
33	pyrophosphatase*	irreversible rxn
34	pyruvate carboxylase	irreversible rxn
35	serine hydroxymethyltransferase	reversible rxn
36	succinate dehydrogenase	reversible rxn
37	sulfate/sulfide reduction	irreversible rxn
38	transaminase*	reversible rxn
39	valine transaminase	reversible rxn

5.3.1.2. Constraints on Chemical Species

As in the astaxanthin study, most of the chemical species are cellular intermediates and are thus assumed not to be accumulated or depleted in an overall fermentation reaction. Additionally, the constraints for AMP, GTP, GDP and PPi are set to prevent accumulation of these species, which will leave the energy requirements of the system expressed in terms of ATP and ADP. Since the first effort is to simply determine the carbon limited conversion, no constraints are initially set on the redox species NAD(P)H and $FADH_2$.

It is not possible before hand to determine whether water and protons will be substrates or products, so no constraint is applied to these species.

For this analysis, glucose and phenylacetic acid are assumed to be the only carbon source, with the optimization effort being directed toward minimizing the amount of glucose needed to produce one mole of penicillin. Sulfate and ammonia are available as substrates. Permitted products other than penicillin include acetate, sulfide, CO_2, α-aminoadipate, and the amino acids cysteine, glycine, serine, and valine.

A point to note is that for this first analysis, α-aminoadipate is allowed to cycle through the reaction network. That is, α-aminoadipate

produced in the final step of penicillin synthesis (5.30) is permitted to reenter the process at reaction 5.24.

The complete set of constraints is shown in Table 5.2.

Table 5.2. Constraints on Chemical Species

1	α-AAA.cys	Not accumulated
2	α-acetolactate	Not accumulated
3	acetyl-CoA	Not accumulated
4	ADP	No constraint
5	α-aminoadipate	Product
6	ammonia	Substrate
7	AMP	Not accumulated
8	ATP	No constraint
9	biomass	Product
10	cis-aconitate	Not accumulated
11	citrate	Not accumulated
12	CO_2	Product
13	CoA	Not accumulated
14	cysteine	Product
15	FAD	No constraint
16	$FADH_2$	No constraint
17	Fd(ox)	Not accumulated
18	Fd(red)	Not accumulated
19	fumarate	Not accumulated
20	GDP	Not accumulated
21	glucose	Maximize
22	glycine	Product
23	glyoxalate	Not accumulated
24	GTP	Not accumulated
25	H_2	Product
26	H_2O	No constraint
27	homocitrate	Not accumulated
28	α,β–dihydroxyisovalerate	Not accumulated
29	isocitrate	Not accumulated
30	isopenicillin.N	Not accumulated

31	α-ketoadipate	Not accumulated
32	α-ketoglutarate	Not accumulated
33	α-ketoisovalerate	Not accumulated
34	malate	Not accumulated
35	NAD^+	No constraint
36	NADH	No constraint
37	$NADP^+$	No constraint
38	NADPH	No constraint
39	O_2	Substrate
40	oxaloacetate	Not accumulated
41	penicillin	Product-w/coeff = 1
42	phenylacetic.acid	Substrate
43	phenylacetyl-CoA	Not accumulated
44	Pi	No constraint
45	PPi	Not accumulated
46	proton	No constraint
47	pyruvate	Not accumulated
48	serine	Product
49	succinate	Not accumulated
50	sulfate	Substrate
51	sulfide	Product
52	tripeptide.1	Not accumulated
53	tripeptide.2	Not accumulated
54	tripeptide.3	Not accumulated
55	valine	Product

In this case, to determine directly how much glucose is required for production of one mole of penicillin, the coefficient of penicillin is fixed. It should be noted that although the optimization goal is to minimize the amount of glucose used to produce penicillin, the constraint on glucose states maximize. This is because the coefficient for glucose as a substrate will be negative. Reducing the amount of glucose consumed will reduce the absolute value of the coefficient, but increase its actual value. Consequently, to minimize the absolute value of a substrate, the constraint for the optimization routine will call for its actual value to be maximized.

5.3.1.3. Results

Optimizing the system of reactions to discern the minimum amount of glucose required for penicillin synthesis under the constraints described above results in a net reaction of:

$$\begin{aligned} &1.5\ \text{glucose} + 1\ \text{phenylacetic acid} + 2\ \text{ammonia} + 1\ \text{sulfate} \\ &+ 6\ \text{ATP} + 3\ \text{NADH} + 2\ \text{FAD} = \\ &\qquad 1\ \text{penicillin} + 1\ CO_2 + 6\ \text{ADP} + 2\ FADH_2 + 3\ NAD^+ \\ &\qquad 3\ H_2O + 6\ \text{Pi} + 3\ H^+ \end{aligned} \tag{5.40}$$

This result is much the same as that obtained by Cooney and Acevedo (1977). Their result for the same conditions was:

$$1.5 \text{ glucose} + 1 \text{ phenylacetic acid} + 2 \text{ ammonia} + 1 H_2SO_4 + 5 \text{ ATP} + 2 \text{ NADH} = 1 \text{ penicillin} + 1 CO_2 + 2 FADH_2 + 7 H_2O \quad (5.41)$$

In contrast, the maximum yield of penicillin on glucose considering only macroscopic balances without regard to the reaction system structure is:

$$1\ 1/3 \text{ glucose} + 1 \text{ phenylacetic acid} + 2 \text{ ammonia} + 1 H_2SO_4 + 3 H_2 = 1 \text{ penicillin} + 10 H_2O \quad (5.42)$$

The net reaction 5.40 consumes 6 ATP and 3 NADH while that of Cooney and Acevedo consumes only 5 ATP and 2 NADH. The difference is that in the reaction scheme behind equation 5.40, it is assumed that the synthesis of the amino acid valine proceeds directly from 2 pyruvate molecules condensing to form acetolactate. Cooney and Acevedo apparently assume that one of the pyruvates would first be decarboxlyated to acetate and then be used in acetolactate formation. However, since acetate must first be converted to pyruvate for the formation of acetolactate to occur (Demain, 1966), acetate formation would introduce a number of seemingly

unnecessary steps into the pathway. Consequently, acetate formation was not included in this analysis for valine synthesis. There is some difference in the balance of water arising from the fact that the sulfate ion was used as the sulfur source in this analysis while Cooney and Acevedo used H_2SO_4 as the sulfur source.

The net reaction shown in equation 5.40 represents the linear combination of the selected reactions on the pathway with rates as shown in Table 5.3. As noted in Chapter 4, these rates are relative rates, not absolute rates. The magnitudes indicate the relative rate of each reaction in relation to another reaction in the pathway. Those reactions shown in Table 5.1, but *not in* Table 5.3, have rates of zero for this solution.

Table 5.3. Optimized Relative Reaction Rates for Carbon Limited Penicillin Production

Rate	Reaction
1	5.24 - irreversible rxn
1	5.25 - irreversible rxn
1	5.26 - reversible rxn
1	5.27 - reversible rxn
1	5.28 - reversible rxn
1	5.29 - irreversible rxn
1	5.30 - reversible rxn
1	acetolactate mutase/reductase* - reversible rxn
1	acetolactate synthase - reversible rxn
1	aconitase(1) - reversible rxn
1	aconitase(2) - reversible rxn
3	AMP kinase - reversible rxn
1	citrate synthetase - irreversible rxn
1	cysteine synthase - reversible rxn
1	dihydroxyacid dehydratase - reversible rxn
1	fumarase - reversible rxn
1	isocitrate lyase - irreversible rxn
1	malate dehydrogenase - reversible rxn
1.5	Overall Embden-Meyerhof pathway - irreversible rxn
1	Overall pyruvate dehydrogenase complex - irreversible rxn
2	pyrophosphatase* - irreversible rxn
1	serine hydroxymethyltransferase - reversible rxn
1	succinate dehydrogenase - reversible rxn
1	sulfate/sulfide reduction - irreversible rxn
1	transaminase* - reversible rxn
1	valine transaminase - reversible rxn

5.3.2. Reduction and Energy Limitations

What should be examined in light of equation 5.40 is the actual yield of penicillin when the requirements for energy and reducing power are taken into account. The optimized yield of equation 5.40 indicates that energy, in the form of six ATP, and reducing and oxidizing power, in the form of three NADH and two FAD, are needed to drive the optimized reaction. The system should be analyzed to discern what effect generating these species has on the final yield of desired product.

5.3.2.1. Constraints on Reactions

No additional constraints are required on the reactions of the original network. It should be noted that, in accordance with Cooney and Acevedo's 1977 analysis of this pathway, $FADH_2$ is oxidized only by electron level phosphorylation or in a reaction for which $FADH_2$ is a cofactor. $FADH_2$ does not reduce NADH in an oxidoreductase reaction. In electron level phosphorylation, the oxidation of one molecule of $FADH_2$ is assumed to yield two ATP molecules.

5.3.2.2. Constraints on Chemical Species.

There are essentially two additional types of constraints added to those in Table 5.2. These constraints involve energy carrying species and redox species.

The first additional constraint is for the energy carrying species ATP and ADP. The production of penicillin is expected to be carried out with glucose providing both the carbon for the synthesis of the product (with the exception of the phenylacetate moiety provided in the feed stock) and for the production of energy. Thus no source of energy in the form of ATP is expected other than that provided by the metabolic oxidation of glucose. A constraint to reflect this is that ADP is expected to be a substrate in the net reaction, with ATP being a product.

The second constraint covers the various redox species NADH, NADPH, $FADH_2$, and their oxidized forms. Since these species are present in only catalytic concentrations in the cell, they are not expected to participate in the net reaction in stoichiometric quantities. To reflect this assumption, each of the species is assigned a coefficient of zero in the net reaction by constraining the species to not be accumulated. In practice only one form, either oxidized or reduced, of each species need be constrained. The coefficient for the second form of the species is not

independent; it will always carry a coefficient of the same magnitude and opposite sign.

5.3.2.3. Results

With the constraint described in section 5.3.2.2 added to the constraints of section 5.3.1.2 the optimization routine yield a solution for the net reaction as shown below:

$$2\ \text{glucose} + 1\ \text{phenylacetic.acid} + 2\ \text{ammonia} + 4\ O_2 + 1\ \text{sulfate} = 1\ \text{penicillin} + 4\ CO_2 + 11\ H_2O \qquad (5.43)$$

In their analysis, Cooney and Acevedo identified a need for a total of only 1.8 moles of glucose per mole of penicillin. Part of the ten percent difference in substrate requirements between the two analyses arises from the additional mole of ATP identified here as required in penicillin synthesis. Another part of the difference is that Cooney and Acevedo assume a yield of six ATP generated by substrate level phosphorylation in glucose oxidation. Here that net yield is reduced to four ATP generated per mole of glucose by substrate level phosphorylation, since six ATP are assumed to be generated, but two are consumed in the overall process.

Table 5.4 show the reaction rates contributing to the solution described by equation 5.43.

Table 5.4. Optimized Relative Reaction Rates for Energy and Reduction Limited Penicillin Production

Rate	Reaction
1	5.24 - irreversible rxn
1	5.25 - irreversible rxn
1	5.26 - reversible rxn
1	5.27 - reversible rxn
1	5.28 - reversible rxn
1	5.29 - irreversible rxn
1	5.30 - reversible rxn
1	acetolactate mutase/reductase* - reversible rxn
1	acetolactate synthase - reversible rxn
1	aconitase(1) - reversible rxn
1	aconitase(2) - reversible rxn
3	AMP kinase - reversible rxn
1	citrate synthetase - irreversible rxn
1	cysteine synthase - reversible rxn
1	dihydroxyacid dehydratase - reversible rxn
1	fumarase - reversible rxn
1	GTP-diphosphate kinase - reversible rxn
1	isocitrate lyase - irreversible rxn
1	malate dehydrogenase - reversible rxn
2	Overall Embden-Meyerhof pathway - irreversible rxn
2	Overall pyruvate dehydrogenase complex - irreversible rxn
1	Overall TCA Cycle - irreversible rxn
3	oxidation of $FADH_2$ (net) - irreversible rxn
2	Oxidative Phosphorylation (mit/mal-asp NADH) - irreversible
2	pyrophosphatase* - irreversible rxn
1	serine hydroxymethyltransferase - reversible rxn
1	succinate dehydrogenase - reversible rxn
1	sulfate/sulfide reduction - irreversible rxn

1 transaminase* - reversible rxn

1 valine transaminase - reversible rxn

5.4. Summary

This case study examined the potential yield available for fermentative production of penicillin from a glucose substrate. It demonstrated the use of linear optimization techniques to rapidly determine optimum relative reaction rates and corresponding stoichiometric coefficients for a reaction network consisting of 39 enzymatic reactions and 55 chemical species. The potential maximum yield of penicillin (benzylpenicillin or penicillin G) was found to be one mole of penicillin produced per two moles of glucose consumed with other reactants and products consumed and produced as indicated by equation 5.43. This result corresponds closely with that previously reported (Cooney and Acevedo, 1977). The relative rates of the optimized network's constituent reactions have not previously been reported and are given here in Tables 5.3 and 5.4. The knowledge of these rate estimates may be useful in manipulating both the growth medium and the metabolism of *Penicillium chrysogenum* in order to increase product yield.

Chapter 6. Case Study III: Aspects of 1,3-Propanediol Biosynthesis

6.1. Overview of the Process

The biological production of 1,3-propanediol is of potential commercial interest. 1,3-propanediol is used in the synthesis of specialty chemicals and polymers, as well as being used as a solvent. Since 1,3-propanediol is presently produced commercially through organic synthesis from petrochemical precursors, the biological production of this species would provide for the possibility of using more readily renewable feedstocks for its production.

The pathway for the production of 1,3-propanediol (1,3-PD) outlined in the following sections involves the combination of the short pathway for direct 1,3-PD synthesis from *Klebsiella pneumoniae* with some metabolic pathways of *Escherichia coli*. This particular combination of organism and pathway was chosen because it allows for the study of the production of a useful, novel product, 1,3-propanediol, in an organism, *E. coli*, for which the metabolism and genetics are well defined.

Klebsiella pneumoniae produces 1,3-PD from glycerol when grown anaerobically, while *E. coli* does not grow on glycerol under anaerobic conditions. The first step in producing 1,3-PD using *E. coli* is to clone the

genes for the reactions in 1,3-PD path into *E. coli* , and to have them expressed so that *E. coli* could produce 1,3-PD on a glycerol substrate.

Since glycerol is more costly than glucose, the next step would be to enable *E. coli* to produce 1,3-PD when grown on glucose or a similarly available sugar substrate. This seems to be a reasonable strategy as there is a common intermediate in both the glycerol to 1,3-PD pathway and in the Embden-Meyerhof glycolytic pathway. The common intermediate is dihydroxyacetone phosphate (DHAP). The general approach would be to divert DHAP from the glycolytic pathway into the pathway for 1,3-PD production.

This analysis concentrates on what the maximal yields of 1,3-PD by *Escherichia coli* on glucose are under several conditions. As such, the analysis deals with a notional biochemical reaction network, a network which has not yet been fully realized. It is hoped that the analysis will provide insight into what issues and problems might be important in developing the system.

The path for the production of propanediol by *Klebsiella pneumoniae* is itself rather short, involving only a few enzymatic reaction steps to modify glycerol to the desired product (Forage and Lin, 1982; Johnson et al., 1984; Johnson et al., 1985; Johnson and Lin, 1987).

Much of the *E. coli* cell metabolic pathways are modeled in some detail (Gottschalk, 1985), because the propanediol synthesis consumes both energy, in the form of ATP, and reducing power, in the form of NADH and $FADH_2$. The metabolic pathways are consequently modeled to investigate what effects different modes for the generation of energy and reducing power might have on the potential yields of propanediol. Thus the pathways modeled include those for aerobic as well as anaerobic metabolism.

6.2. Pathways for 1,3-Propanediol Biosynthesis

The pathway network outlined here is a modification of the propanediol pathway outlined by I-teh Tong, whose inestimable contributions to this work are greatly appreciated.

6.2.1. Pyruvate Formation

The pathway for the formation of pyruvate from a glucose feedstock by *E. coli* is outlined in Figure 6.1. This path includes both the Embden-Meyerhof pathway and an additional path involving the formation of methylglyoxal.

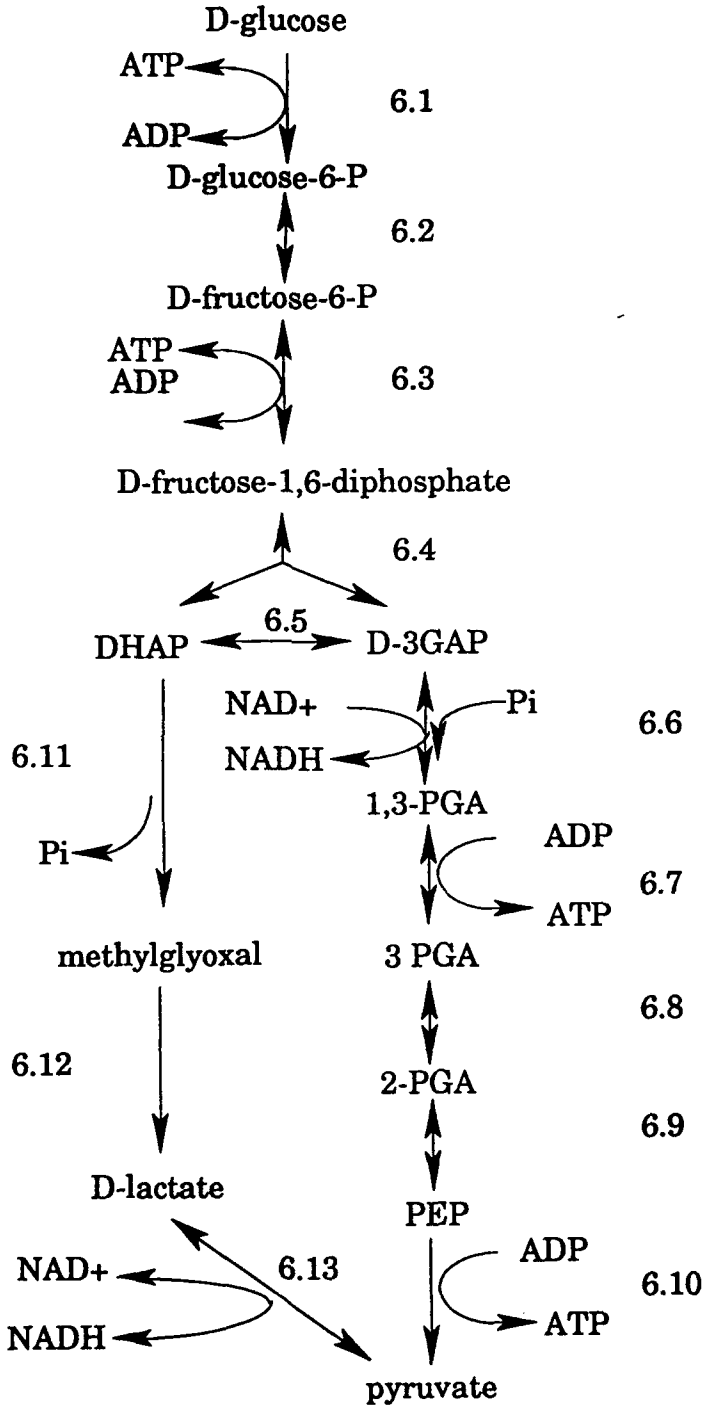


Figure 6.1. Pyruvate Formation

Key to reactions in Figure 6.1.

6.1, hexokinase; 6.2, glucose-6-phosphate isomerase;

6.3, 6-phosphofructokinase; 6.4, fructose bisphosphate aldolase;

6.5, triosephosphate isomerase;

6.6, glyceraldehyde-3-phosphate dehydrogenase;

6.7, phosphoglycerate kinase; 6.8, phosphoglycerate mutase; 6.9, enolase;

6.10, pyruvate kinase; 6.11, methylglyoxal synthetase; 6.12, glyoxalase;

6.13, D-lactate dehydrogenase

Abbreviations in Figure 6.1.

2 PGA: 2-phosphoglycerate

3 PGA: 3-phosphoglycerate

1,3-PGA: 1,3-bisphosphoglycerate

D-3GAP: D-glyceraldehyde-3-phosphate

DHAP: dihydroxyacetone phosphate

PEP: phosphoenolpyruvate

6.2.2. Reduction to 1,3-Propanediol

The pathway for the synthesis of 1,3-propanediol from dihydroxyacetone phosphate by *K. pneumoniae* is depicted in Figure 6.2 (Forage and Lin, 1982; Johnson et al., 1984; Johnson et al., 1985; Johnson and Lin, 1987).

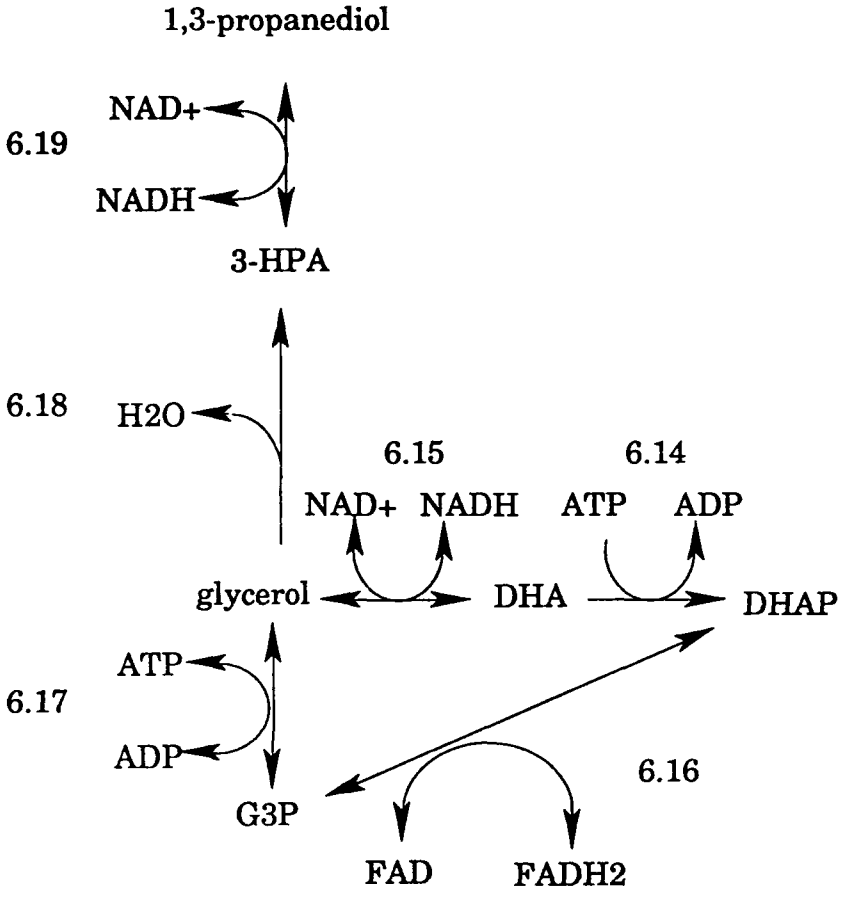


Figure 6.2. 1,3-Propanediol Synthesis Pathway

Key to reactions in Figure 6.2.

6.14, dihydroxyacetone kinase*; 6.15, glycerol dehydrogenase;

6.16, glycerol phosphate dehydrogenase; 6.17, glycerol kinase;

6.18, glycerol dehydratase; 6.19, 1,3-propanediol dehydrogenase

Abbreviations in Figure 6.2.

3-HPA: 3-hydroxypropionaldehyde; DHA: dihydroxyacetone

DHAP: dihydroxyacetone phosphate; G3P: glycerol-3-phosphate

6.2.3. Pyruvate Dehydrogenase and Partial Tricarboxylic Acid Cycle

The major metabolic pathways used by *E. coli* for energy production are depicted in Figure 6.3. These paths are important in the production of carbon species which can enter into other metabolic processes in the cell or be further modified (for example, being further reduced) and appear as significant extracellular products.

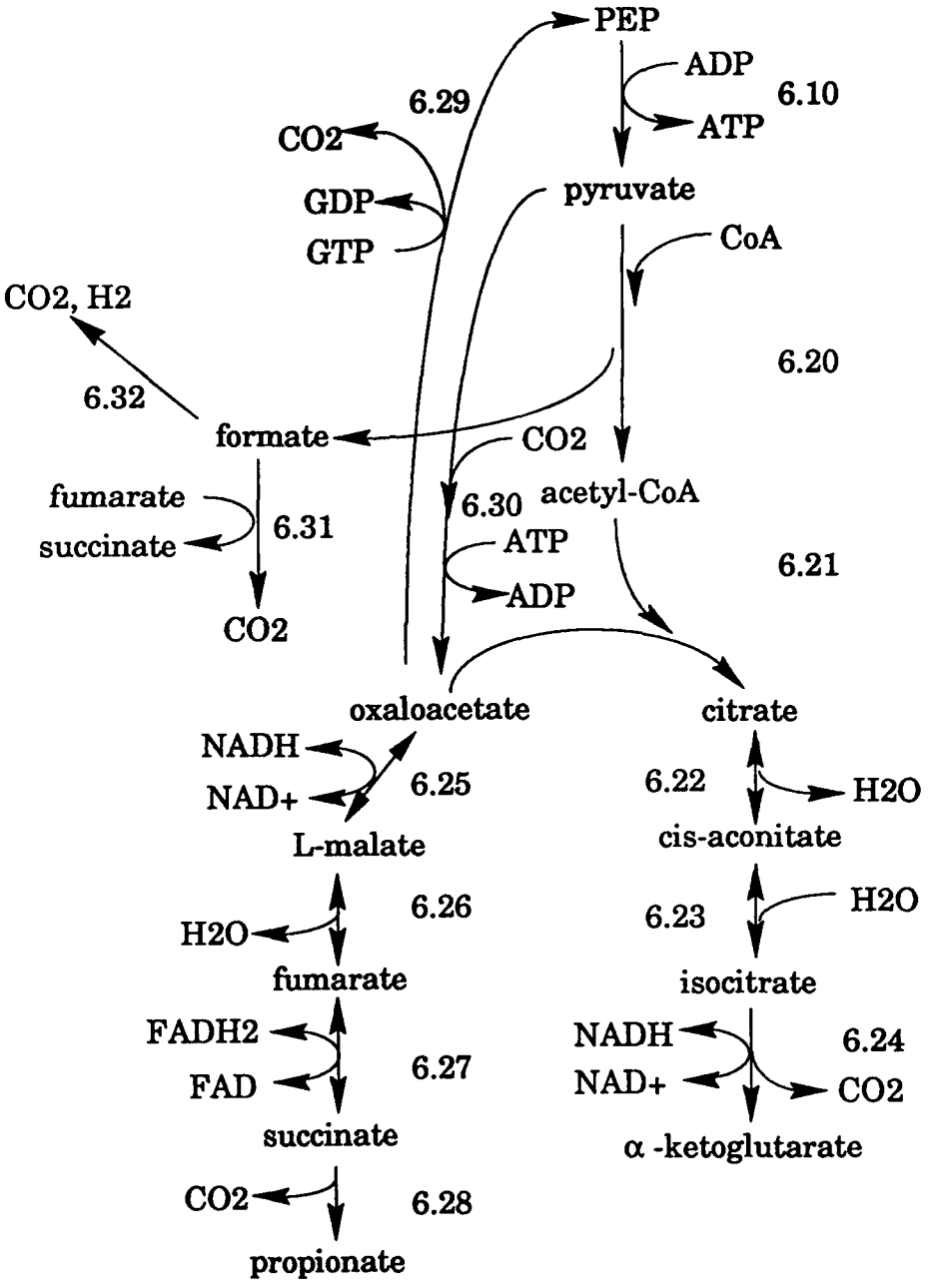


Figure 6.3. Partial Tricarboxylic Acid Cycle

Key to reactions in Figure 6.3.

6.10, pyruvate kinase; 6.20, pyruvate dehydrogenase;

6.21, citrate synthase; 6.22, aconitase (1)*; 6.23, aconitase (2)*;

6.24, isocitrate dehydrogenase; 6.25, malate dehydrogenase;

6.26, fumarase; 6.27, succinate dehydrogenase;

6.28, succinate decarboxylase; 6.29, phosphoenolpyruvate carboxykinase;

6.30, pyruvate carboxylase; 6.31, formate-hydrogen lyase(1);

6.32, formate-hydrogen lyase(2)

6.2.4. Reduced Metabolites

One set of pathways in *E. coli* for the elimination of excess reducing energy from the cell is depicted in Figure 6.4. This process involves producing carbon products which are more reduced than the original glucose feedstock which serve as electron sinks for the cell. Use of these sinks permit the cell to metabolize the feedstock in the absence of external electron acceptors.

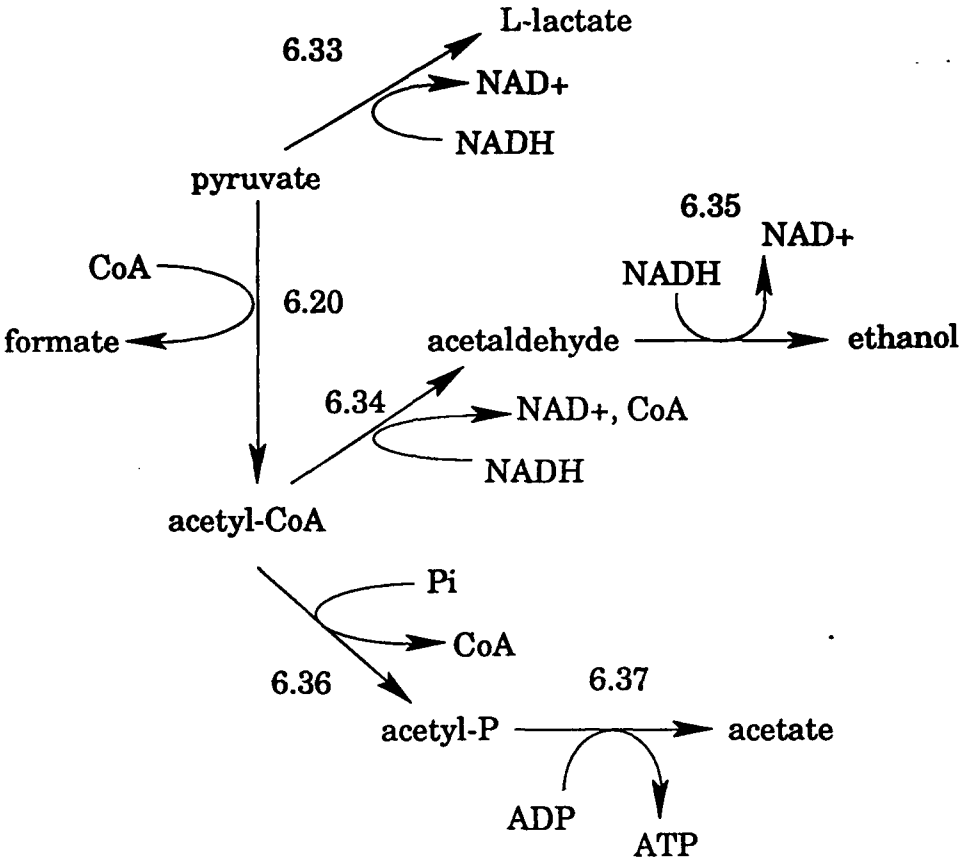


Figure 6.4. Reduced Metabolic Products

Key to reactions in Figure 6.4.

6.20, pyruvate dehydrogenase; 6.33, L-lactate dehydrogenase;

6.34, acetaldehyde dehydrogenase; 6.35, ethanol dehydrogenase;

6.36, phosphotransacetylase; 6.37, acetate kinase

6.2.5. Pentose Phosphate Pathway

In Figure 6.5, the pentose phosphate pathway is illustrated. One of the potentially important aspects of this pathway in *E. coli* is that it permits the cell to produce reducing power in the form of NADPH. The significance of the path in this analysis is that it might provide a method for producing the reducing power needed for the synthesis of 1,3-propanediol.

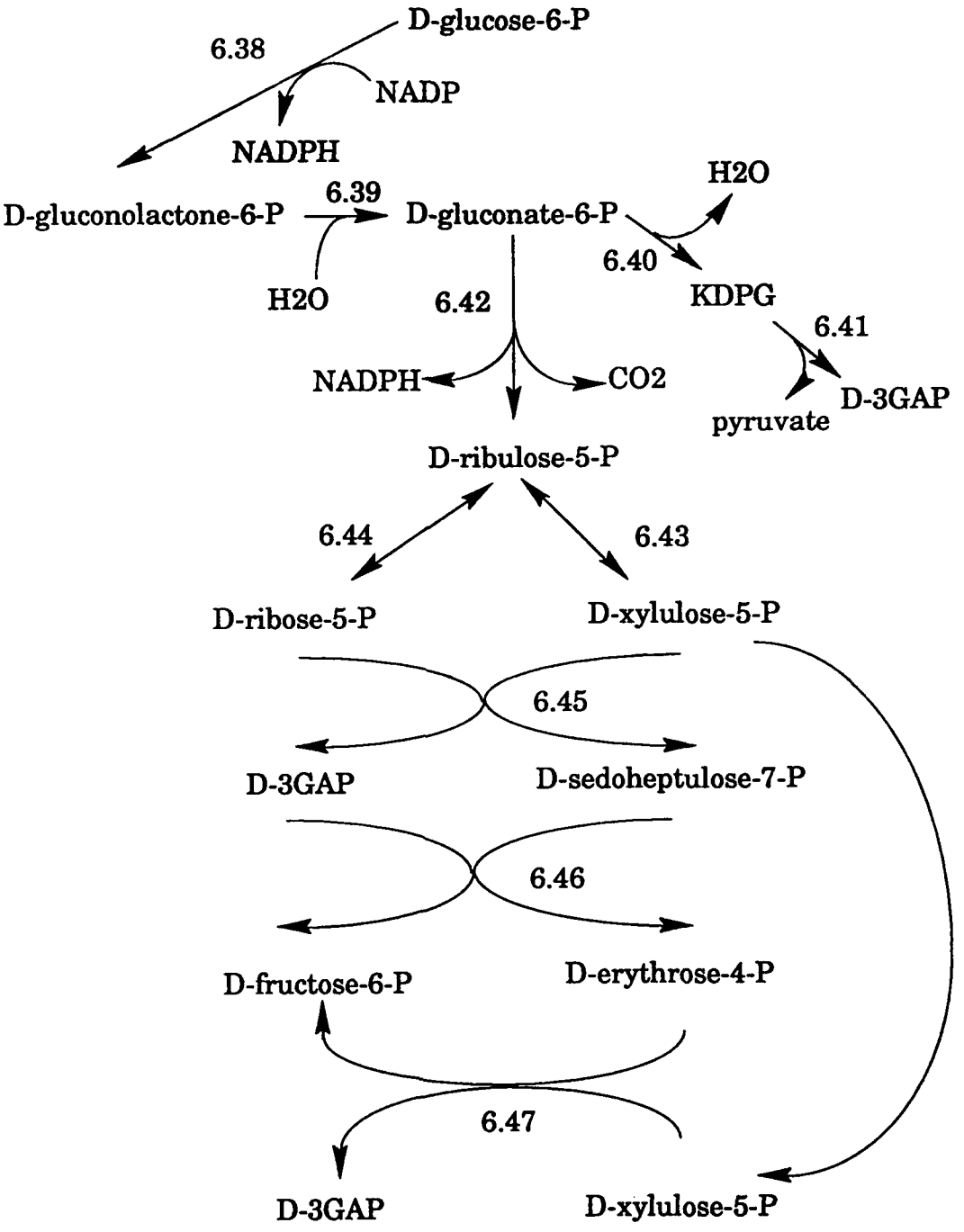


Figure 6.5. Pentose Phosphate Pathway

Key to reactions in Figure 6.5.

6.38, glucose-6-phosphate dehydrogenase; 6.39, lactonase;

6.40, gluconate dehydratase; 6.41, ketodeoxyphosphogluconate aldolase;

6.42, 6-phosphogluconate dehydrogenase;

6.43, phosphopentose epimerase; 6.44, phosphopentose isomerase;

6.45, transketolase (5/5-3/7)*; 6.46, transaldolase (7/3-4/6)*;

6.47, transketolase (5/4-3/6)*

Abbreviations in Figure 6.5.

D-3GAP: glyceraldehyde-3-phosphate

KDPG: 2-keto-3-deoxygluconate-6-phosphate

6.2.6. Other Reactions

As in the previous two case studies, several additional enzymatic reactions are included to account for conversions between various chemical species, as well as for the production of biomass. (6.49 and further)

ATP and GTP may be interconverted through the action of GTP diphosphate kinase:

$$ATP + GDP = ADP + GTP \quad (6.48)$$

If it is assumed that the intracellular redox species may be interconverted by oxidoreductases, the following reactions should be included:

$$FADH_2 + NAD^+ = FAD + NADH + H^+ \quad (6.49)$$

$$NADP^+ + NADH = NADPH + NAD^+ \quad (6.50)$$

Again, a generalized formula for the production of biomass is used which is based on average biomass formulations and on and intermediate level of energy requirement for biomass synthesis. This formula may indeed not precisely reflect the composition of *E. coli* under the desired production conditions; however, the goal of the optimization is to produce

1,3-PD with as little cell growth as possible. Thus the approximation will serve unless a large degree of the substrate goes to cell growth instead of product formation.

$$4/3 \text{ glucose} + \text{ammonia} + \text{NADH} + 19.8\ \text{ATP} + 15.8\ H_2O = \text{biomass} + NAD^+ + 19.8\ \text{ADP} + 19.8\ \text{Pi} + 18.8\ H^+ \quad (6.51)$$

6.3. An Analysis of Potential Yields of 1,3-Propanediol

6.3.1. Carbon Limited Biosynthesis

The first analysis of the biosynthetic production of propanediol is to consider what limits on conversion are imposed simply by the topology of the biochemical pathway in processing the carbon source. In this study, glucose, a common growth substrate, is assumed to be the carbon source.

6.3.1.1. Constraints on reactions

The reversibilities and irreversibilities of the reactions in the pathway are indicated in the figures by the directions or bidirections of the arrows on the reaction paths. Table 6.1 provides a summary of the constraints these reaction reversibilities impose on the system.

Table 6.1. Reaction Constraints

1	1,3-propanediol dehydrogenase	reversible rxn
2	6-phosphofructokinase	reversible rxn
3	6-phosphogluconate dehydrogenase	irreversible rxn
4	acetaldehyde dehydrogenase	irreversible rxn
5	acetate kinase	irreversible rxn
6	aconitase(1)	reversible rxn
7	aconitase(2)	reversible rxn
8	biomass production	irreversible rxn
9	citrate synthetase	irreversible rxn
10	D-lactate dehydrogenase	reversible rxn
11	dihydroxyacetone kinase	irreversible rxn
12	enolase	reversible rxn
13	ethanol dehydrogenase	irreversible rxn
14	formate-hydrogen lyase(1)	irreversible rxn
15	formate-hydrogen lyase(2)	irreversible rxn
16	fructose-bisphosphate aldolase	reversible rxn
17	fumarase	reversible rxn
18	gluconate dehydratase *	irreversible rxn
19	glucose-6-phosphate dehydrogenase	irreversible rxn
20	glucose-6-phosphate isomerase	reversible rxn
21	glyceraldehyde-3-phosphate dehydrogenase	reversible rxn
22	glycerol dehydratase	irreversible rxn
23	glycerol dehydrogenase	reversible rxn
24	glycerol kinase	reversible rxn
25	glycerol-phosphate dehydrogenase	reversible rxn
26	glyoxalase	irreversible rxn
27	GTP-diphosphate kinase	reversible rxn
28	hexokinase	irreversible rxn
29	isocitrate dehydrogenase	irreversible rxn
30	ketodeoxyphosphogluconate aldolase*	irreversible rxn

31	L-lactate dehydrogenase	irreversible rxn
32	lactonase	irreversible rxn
33	malate dehydrogenase	reversible rxn
34	methylglyoxal synthase	irreversible rxn
35	NAD(P)H oxidoreductase*	reversible rxn
36	NAD/FADH2 oxidoreductase*	reversible rxn
37	overall pyruvate dehydrogenase complex*	irreversible rxn
38	phosphoglycerate kinase	reversible rxn
39	phosphoglycerate mutase	reversible rxn
40	phosphoenolpyruvate carboxykinase	irreversible rxn
41	phosphopentose epimerase	reversible rxn
42	phosphopentose isomerase	reversible rxn
43	phosphotransacetylase	reversible rxn
44	pyruvate carboxylase	irreversible rxn
45	pyruvate kinase	irreversible rxn
46	succinate decarboxylase*	irreversible rxn
47	succinate dehydrogenase	reversible rxn
48	transaldolase (7/3-4/6)	reversible rxn
49	transketolase (5/4-3/6)	reversible rxn
50	transketolase (5/5-7/3)	reversible rxn
51	triose phosphate isomerase	reversible rxn

6.3.1.2. Constraints on Chemical Species

To determine what the carbon limited conversion yield is, the first assumption is that glucose is the only carbon source available. Also available is nitrogen in the form of ammonia. The various major carbon products which experiments indicate may be expected from the fermentation process are: carbon dioxide, acetate, biomass, ethanol, formic acid, D-lactate, L-lactate, and succinate.

No constraints are placed on the water, protons, or the reduction/oxidation species, nor are any constraints placed on the energy carrying species.

The remainder of the chemical species are assumed to be metabolic intermediates which do not participate in the net system reaction in stoichiometric quantities. The constraint for those species is set such that they are not accumulated or consumed.

For this analysis, the optimization goal is to minimize the amount of glucose required to form one mole of 1,3-propanediol. The full set of constraints on the chemical species involved in the reaction network is enumerated in Table 6.2. It should be noted that indicated constraint for glucose is to maximize its coefficient. The constraint is stated in this way, because glucose is to be a substrate. Since substrates have negative

stoichiometric coefficients, reducing the amount of glucose will increase the actual value of its coefficient. The linear optimization routine operates on the actual values of the coefficients, so the constraint to minimize the amount of glucose used is in reality a constraint to maximize the actual value of its coefficient.

Table 6.2. Constraints on Chemical Species

1	acetaldehyde	Not Accumulated
2	acetate	Product
3	acetyl-CoA	Not Accumulated
4	acetyl-P	Not Accumulated
5	ADP	No constraint
6	ammonia	Substrate
7	ATP	No constraint
8	biomass	Product
9	1,3-bisphosphoglycerate	Not Accumulated
10	cis-aconitate	Not Accumulated
11	citrate	Not Accumulated
12	CO_2	Product
13	CoA	Not Accumulated
14	dihydroxyacetone	Accumulated
15	dihydroxyacetone.phosphate	Not Accumulated
16	erythrose-4-phosphate	Not Accumulated
17	ethanol	Product
18	FAD	No constraint
19	$FADH_2$	No constraint
20	formate	Product
21	fructose.1,6-bisphosphate	Not Accumulated
22	fructose.6-phosphate	Not Accumulated
23	fumarate	Not Accumulated
24	GDP	Not Accumulated
25	D-gluconate-6-phosphate	Not Accumulated
26	D-gluconolactone-6-phosphate	Not Accumulated
27	glucose	Maximize
28	glucose.6-phosphate	Not Accumulated
29	glyceraldehyde.3-phosphate	Not Accumulated
30	glycerol	Not Accumulated

31	glycerol-3-phosphate	Not Accumulated
32	GTP	Not Accumulated
33	H_2	Product
34	H_2O	No constraint
35	3-hydroxypropionaldehyde	Not Accumulated
36	isocitrate	Not Accumulated
37	2-keto-3-deoxygluconate-6-phosphate	Not Accumulated
38	α-ketoglutarate	Not Accumulated
39	D-lactate	Product
40	L-lactate	Product
41	malate	Not Accumulated
42	methylglyoxal	Not Accumulated
43	NAD^+	No constraint
44	NADH	No constraint
45	$NADP^+$	No constraint
46	NADPH	No constraint
47	oxaloacetate	Not Accumulated
48	phosphoenolpyruvate	Not Accumulated
49	2-phosphoglycerate	Not Accumulated
50	3-phosphoglycerate	Not Accumulated
51	Pi	No constraint
52	1,3-propanediol	Product-w/coeff = 1
53	propionate	Not Accumulated
54	proton	No constraint
55	pyruvate	Not Accumulated
56	ribose-5-phosphate	Not Accumulated
57	ribulose-5-phosphate	Not Accumulated
58	sedoheptulose-7-phosphate	Not Accumulated
59	succinate	Product
60	xylulose-5-phosphate	Not Accumulated

6.3.1.3. Results

When the reaction system is solved to minimize the consumption of glucose for production of propanediol (i.e. maximizing the yield of propanediol on glucose) subject to the constraints outlined in Tables 6.1 and 6.2, the following net reaction equation is obtained:

$$0.5\ \text{glucose} + 1\ FADH_2 + 1\ NADH + 1\ \text{proton} = 1\ \text{1,3-propanediol} + 1\ FAD + 1\ NAD^+ + 1\ H_2O \quad (6.53)$$

The net reaction is based on a reaction network with reaction rates as depicted in Table 6.3.

Table 6.3. Optimized Rates for Carbon Limited Production

Rate	Reaction
1	1,3-propanediol dehydrogenase - reversible rxn
0.5	6-phosphofructokinase - reversible rxn
0.5	fructose-bisphosphate aldolase - reversible rxn
0.5	glucose-6-phosphate isomerase - reversible rxn
1	glycerol dehydratase - irreversible rxn
-1	glycerol kinase - reversible rxn
-1	glycerol-phosphate dehydrogenase - reversible rxn
0.5	hexokinase - irreversible rxn
-0.5	triose phosphate isomerase - reversible rxn

What can readily be seen from this analysis is that the structure of the reaction network is such that it might be possible to convert all of the carbon in the feedstock to the desired product if some external source of reducing power were available to the cell to bring the glucose feed to the reduction level of propanediol.

6.3.2. Reducing Power Limits to Conversion Yield

For the next step in the analysis, the effects of the requirements for reducing power are examined.

6.3.2.1. Constraints on Reactions

No additional constraints on the reactions are required beyond those derived and listed in Table 6.1.

6.3.2.2. Constraints on Chemical Species

The additional constraints required to reflect the lack of external reducing power are introduced by constraining the various reducing and oxidizing species in the cell to be not accumulated or consumed. These species include NADH, NADPH, $FADH_2$, and ferredoxin in both their reduced and oxidized forms.

6.3.2.3. Results

When the reaction system is optimized to give the maximum possible yield of propanediol on glucose, the net reaction which results is:

$$0.6667 \text{ glucose} = 1 \text{ 1,3-propanediol} + 1\ CO_2 \qquad (6.54)$$

The underlying reaction network which produces this result is delineated in Table 6.4.

Table 6.4. Optimized Reaction Rates for Reduction Limited Synthesis

Rate	Reaction
1	1,3-propanediol dehydrogenase - reversible rxn
0.3333	6-phosphofructokinase - reversible rxn
1	6-phosphogluconate dehydrogenase - irreversible rxn
0.3333	fructose-bisphosphate aldolase - reversible rxn
1	glucose-6-phosphate dehydrogenase - irreversible rxn
-0.3333	glucose-6-phosphate isomerase - reversible rxn
1	glycerol dehydratase - irreversible rxn
-1	glycerol kinase - reversible rxn
-1	glycerol-phosphate dehydrogenase - reversible rxn
0.6667	hexokinase - irreversible rxn
1	lactonase - irreversible rxn
2	NAD(P)H oxidoreductase* - reversible rxn
-1	$NAD/FADH_2$ oxidoreductase* - reversible rxn
0.6667	phosphopentose epimerase - reversible rxn
0.3333	phosphopentose isomerase - reversible rxn
0.3333	transaldolase (7/3-4/6) - reversible rxn
0.3333	transketolase (5/4-3/6) - reversible rxn
0.3333	transketolase (5/5-7/3) - reversible rxn
-0.6665	triose phosphate isomerase - reversible rxn

What should be noted about this particular solution is that it involves the use of the pentose phosphate pathway as a sources of reducing power. There is some question as to whether this pathway can actually be made to function at the same time that there is substantial activity in the Embden-Meyerhof path. These two paths would normally not be significantly expressed in the cell concurrently. Clearly, the production of 1,3-propanediol with both pathways expressed would be very efficient. Three-quarters of the carbon introduced as feedstock could potentially be converted to the desired product with the only significant byproduct being carbon dioxide.

What should be considered next is the possible yield and distribution of products if the pentose phosphate pathway cannot be expressed at the same time that the Embden-Meyerhof pathway is assimilating the sugar feed into the cell.

6.3.3. Pathway Limited Yields

In this section, the conversion of glucose to 1,3-propanediol in the absence of tangible expression of the pentose phosphate pathway will be considered.

6.3.3.1. Constraints on Reactions

In order to model the effect of not having the pentose phosphate pathway present, the rates of the set of reactions which comprise that path are set to zero. Within the Pathway Toolbox Program other methods are available to produce the same effect. For instance, new reaction network could be constructed without those reactions, but simply setting their rates to zero is by far the quickest method. The reactions in the pathway are:

6.38, glucose-6-phosphate dehydrogenase; 6.39, lactonase;
6.40, gluconate dehydratase;
6.41, ketodeoxyphosphogluconate aldolase;
6.42, 6-phosphogluconate dehydrogenase;
6.43, phosphopentose epimerase; 6.44, phosphopentose isomerase;
6.45, transketolase (5/5-3/7)*; 6.46, transaldolase (7/3-4/6)*;
6.47, transketolase (5/4-3/6)*

6.3.3.2. Constraints on Chemical Species

The constraints on the coefficients for the various chemical species in the reaction system remain the same as those used in section 6.3.2.

6.3.3.3. Results

After the reaction network with appropriate constraints is processed through the linear optimization routine, the first solution derived is a net reaction for the system of:

$$1 \text{ glucose} = 1 \text{ 1,3-propanediol} + 1 \text{ } CO_2 + 1 \text{ acetate} \qquad (6.55)$$

The reaction rates for this solution are laid out in Table 6.5.

Table 6.5. Optimum Rates with Acetate as Byproduct

Rate	Reaction
1	1,3-propanediol dehydrogenase - reversible rxn
1	6-phosphofructokinase - reversible rxn
1	acetate kinase - irreversible rxn
-1	D-lactate dehydrogenase - reversible rxn
1	formate-hydrogen lyase(1) - irreversible rxn
1	fructose-bisphosphate aldolase - reversible rxn
1	glucose-6-phosphate isomerase - reversible rxn
1	glycerol dehydratase - irreversible rxn
-1	glycerol kinase - reversible rxn
-1	glycerol-phosphate dehydrogenase - reversible rxn
1	glyoxalase - irreversible rxn
1	hexokinase - irreversible rxn
1	methylglyoxal synthase - irreversible rxn
1	overall pyruvate dehydrogenase complex* - irreversible rxn
1	phosphotransacetylase - reversible rxn
1	succinate dehydrogenase - reversible rxn
-1	triose phosphate isomerase - reversible rxn

Since there are reduced products other than acetate that are possible products of the reaction network, the next analysis might be to determine what net reaction is reached if acetate is disallowed as a product. This might be the case if a mutant of *E. coli* was used which could not produce acetate.

The additional constraint of no acetate production results in the system becoming infeasible. That is there is no possible set of rates for the pathway laid out in section 6.2 which will produce 1,3-propanediol from glucose without either utilizing the pentose phosphate shunt or producing acetate as a byproduct. Thus a limiting case has been reached for the analysis of 1,3-propanediol production from glucose.

6.4. Summary

This limiting case provides some guidance for what can be expected in attempting to optimize the process. In the absence of an external electron source or the expression of the pentose phosphate pathway, the best yield that can be expected for the production of 1,3-propanediol from a pathway cloned into *E. coli* is one mole of product per mole of glucose feed. This yield represents only fifty percent of the carbon in feedstock being converted to the desired product.

An important aspect to consider in attempting to reach this optimal yield is the importance of acetate production in getting to it. An uninformed attempt to increase the yield by eliminating acetate production would be counterproductive; the nominally unwanted product is required by the structure of the pathway for the maximal production of 1,3-propanediol from glucose.

An interesting aspect of the use of linear optimization for this analysis is the rapidity with which the outlines of the maximum yield problem can be delineated. After the initial effort has been invested in determining the correct form of the pathway, only three runs are needed to determine the various levels of possible yields. The first two runs determine what the yields are for the carbon limited and the energy and redox limited cases. The third run determines that the the fermentation must either use the pentose phosphate pathway or produce acetate in order for 1,3-propanediol to be produced. In real time, the three runs on a Macintosh II take less than five minutes. An additional advantage of using the program is that at the end of that time, one can be reasonably sure that there are no mathematical errors in the derived results.

Chapter 7. Further Work

This particular work is quite obviously only a small step in the direction of obtaining all the information which linear programming techniques could be used to extract from the stoichiometric structure of a biochemical reaction system. In terms of optimization using stoichiometric data, there are two additional fundamental aspects of stoichiometric modeling to be investigated: more realistic objective functions, and creating a more simple way to put data into and extract data from the computer. The data input and output could be simplified both in terms of mechanical aspects of data entry and representation and in terms of using the machine to make certain types of decisions and calculations without assistance from the user.

7.1. Objective Functions

The objective functions used throughout this work have involved maximizing production of, or minimizing consumption of, a single chemical species. The MINOS optimization package is quite capable of handling any sort of linear objective function. With a moderate amount of programming, the HyperCard front end to MINOS could be modified to permit the introduction of more useful objective functions. These functions should involve more than a single chemical species, and they

should provide for the assignment of differing values to these species. Because of the power available on MINOS as well as other optimization packages, it might prove useful to provide a capability for introducing non-linear objective functions if they represented important aspects of the optimization problem.

7.2. User Interface

7.2.1. User Interface - Representation

The current front end of the program, the portion with which the user directly works, could be vastly improved. The choice of HyperCard as the programming environment allowed for the relatively rapid development of the total package, but it greatly limited the speed and flexibility of the user interface. Programming the user interface in more powerful language such as Pascal or C should permit a far more intuitive and graphical approach to outlining and manipulating the reaction network. Such an approach might be to provide the user the capability to simply draw out the biochemical system to be studied. A functional example of this type of approach is found in MacProject, an optimization program for project management. Of course, a biochemical system presents far more numerous and complex interconnections to be handled than are found in a typical project management problem. This level of relational complexity would present a challenge in finding an

approach that allows the interrelations to be easily represented without grossly overcomplicating the graphic representation. It would hardly be worth the effort to create a graphic interface which makes the pathway no simpler to comprehend than the current approach.

7.2.2. User Interface - Setting Constraints

A vexing problem in setting up the optimization process is that of establishing the various constraints on the reaction rates and on the chemical species stoichiometric coefficients.

Determining what the constraints on the reactions should be involves a good deal of investigation into each reaction, and still often comes down to applying a sparse set of heuristic rules. What is needed is a method for having the machine independently estimate the energy changes involved in each reaction. The user would then later set some arbitrary cutoff for reaction reversibility and/or set a constraint on the energy change for the net reaction of the optimized system.

The constraints on the stoichiometric coefficients of the chemical species in the system present a more difficult problem. While here again a few simple rules of thumb can be applied initially to get a first estimate of the optimized yield, determining which constraints are important and which are unimportant is currently very difficult. In order to solve this

problem, an optimization package is required which will perform sensitivity analyses on the optimization problem it is solving. The data from the sensitivity analysis must then be incorporated into readily comprehensible output which includes not only the net reaction and its associated reaction rates, but also an indication of how sensitive those values are to changes in the constraint set.

Abbreviations

α-AAA-cys-val: α-aminoadipic acid-cysteine-valine

α-AAA-cys: α-aminoadipic acid-cysteine

acetyl-CoA: acetyl coenzyme A

acetyl-P: acetyl phosphate

ADP: adenosine diphosphate

AMP: adenosine monophosphate

ATP: adenosine triphosphate

CO_2: carbon dioxide

CoA: coenzyme A

D-3GAP: glyceraldehyde 3-phosphate

DHA: dihydroxyacetone

DHAP: dihydroxyacetone phosphate

DNA: deoxyribonucleic acid

FAD: flavin adenine dinucleotide (oxidized form)

$FADH_2$: flavin adenine dinucleotide (reduced form)

Fd(ox): ferredoxin (oxidized form)

Fd(red): ferredoxin (reduced form)

G3P: glycerol 3-phosphate

GDP: guanosine diphosphate

GTP: guanosine triphosphate

3-HPA: 3-hydroxypropionaldehyde

NAD^+: nicotinamide adenine dinucleotide (oxidized form)

NADH: nicotinamide adenine dinucleotide (reduced form)

$NADP^+$: nicotinamide adenine dinucleotide phosphate (oxidized form)

$NADPH_2$: nicotinamide adenine dinucleotide phosphate (reduced form)

PEP: phosphoenolpyruvate

1,3 -PD: 1,3-propanediol (trimethylene glycol)

1,3-PGA: 1,3-bisphosphoglycerate

2-PGA: 2-phosphoglycerate

3-PGA: 3-phosphoglycerate

Pi: orthophosphate (HPO_4^{-2})

PPi: pyrophosphate ($HP_2O_7^{-3}$)

References

Absoft Corporation, MacFortran & MacFortran/020, Rochester Hills, Minnesota, 1989.

Aiba, S., and Matsuoka, M., Identification of metabolic model: citrate production from glucose by *Candida lipolytica*, Biotechnol. Bioeng., 21, 1373-1386, (1979).

An, G. H., Schuman, D. B., and Johnson, E. A., Isolation of *Phaffia rhodozyma* mutants with increased astaxanthin content, Appl. Env. Microbiol., 55, 116-124, (1989).

Andrewes, A. G., Phaff, H. J., and Starr, M. P., Carotenoids of *Phaffia rhodozyma*, a red-pigmented fermenting yeast, Phytochemistry, 15, 1003-1007, (1976).

Apple Computer , Inc., HyperCard Script Language Guide, Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1988.

Aris, R. and Mah, R. H. S., Independence of chemical reactions, I&EC Fund., 2, 90-94, (1963).

Aris, R., Elementary Chemical Reactor Analysis, Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1969.

Bauernfeind, J. C., Brubacher, G. B., Kläui, H. M. and Marusich, W. L., Use of carotenoids in: Carotenoids, ed. by Isler, O., Birkhäuser Verlag, Basel, 1971.

Blum, J. J., and Stein, R. B., On the analysis of metabolic networks, 99-125, in Biological Regulation and Development, volume 3A, edited by Goldberger and Yamamoto, Plenum Press, 1982.

Britton, G., Biosynthesis of carotenoids, 262-327, in Chemistry and Biochemistry of Plant Pigments, edited by T. W. Goodwin, Academic Press, New York, 1976.

Britton, G., The Biochemistry of Natural Pigments, Cambridge University Press, Cambridge, 1983.

Cooney, C. L., and Acevedo, F., Theoretical conversion yields for penicillin synthesis, Biotechnol. Bioeng., 19, 1449-1462, (1977).

Cooney, C. L., Conversion yields in penicillin production: theory vs. practice, Proc. Biochem., 31-33, 1979.

Cooney, C. L., Wang, H. Y., and Wang, D. I. C., Computer-aided material balancing for prediction of fermentation parameters, Biotechnol. Bioeng., 19, 55-67, (1977).

Demain, A. L., Biosynthesis of penicillins and cephalosporins, in Biosynthesis of Antibiotics, Volume 1, edited by J. F. Snell, Academic Press, New York, 1966.

Domach, M. M., Leung, S. K., Cahn, R. E., Cocks, G. G. and Shuler, M. L., Computer model for glucose-limited growth of a single cell of *Escherichia coli* B/r-A, Biotechnol. Bioeng., 26, 203-216, (1984).

Erickson, L. E., Minkevich, I. G., and Eroshin, V. K., Application of mass and energy balance regularities in fermentation, Biotechnol. Bioeng., 20, 1595-1621, (1978a).

Erickson, L. E., Selga, S. E., and Viesturs, U. E., Application of mass and energy balance regularities to product formation, Biotechnol. Bioeng., 20, 1623-1638, (1978b).

Fletcher, R., Practical Methods of Optimization, 2d edition, John Wiley & Sons, New York, 1987.

Forage, R. G. and Lin, E. C. C., *dha* System mediating aerobic and anaerobic dissimilation of glycerol in *Klebsiella pneumoniae* NCIB 418, J. Bacteriol., 151, 591-599, (1982).

Georgiou, G., Optimizing the production of recombinant proteins in microorganisms, AIChE J., 34, 1233-1246, (1988).

Gottschalk, G., Bacterial Metabolism, 2d edition, Springer-Verlag, New York , 1985.

Heijnen, J. J., Roels, J. A., and Stouthamer, A. H., Application of balancing methods in modelling the penicillin fermentation, Biotechnol. Bioeng., 21, 2175-2201, (1979).

Hoover, S. R. and Allison, F. E., The growth metabolism of rhizobium, with evidence on the interrelations between respiration and synthesis, J. Biol. Chem., 134, 181-192, (1940).

International Unionof Biochemistry, Nomenclature Committee, Enzyme Nomenclature, Academic Press, Inc., Orlando, Florida, 1984.

Irvine, R. L. and Bryers, J. D., Stoichiometry and kinetics of waste treatment, 757-772, in Comprehensive Biotechnology, volume 4, edited by C. W. Robinson and J. A. Howell, Pergamon Press, New York, 1985.

Johnson, E. A. and Lewis, M. L., Astaxanthin formation by the yeast *Phaffia Rhodozyma*, J. Gen. Microbiol., 115, 173-183, (1979).

Johnson, E. A. and Lin, E. C. C., *Klebsiella pneumoniae* 1,3-propanediol:NAD+ oxidoreductase, J. Bacteriol., 169, 2050-2054, (1987).

Johnson, E. A., Burke, S. K., Forage, R. G., and Lin, E. C. C., Purification and properties of dihydroxyacetone kinase from *Klebsiella pneumoniae* , J. Bacteriol., 160, 55-60, (1984).

Johnson, E. A., Levine, R. L., and Lin, E. C. C., Inactivation of glycerol dehydrogenase of *Klebsiella pneumoniae* and the role of divalent cations, 164, 479-483, (1985).

Johnson, E. A., Lewis, M. J. and Grau, C. R., Pigmentation of egg yolks with astaxanthin from the yeast *Phaffia Rhodozyma*, Poultry Science, 59, 1777-1782, (1980).

Joshi, A. and Palsson, B. O., *Escherichia coli* growth dynamics: a three-pool biochemically based description, Biotechnol. Bioeng., 31, 102-116, (1988).

Joshi, A., Integrated metabolic dynamics in the human red blood cell, Ph.D. Thesis, Department of Chemical Engineering, University of Michigan, 1988.

Kläui, H. M., Industrial and commercial uses of carotenoids, in: Carotenoid Chemistry and Biochemistry, ed. by Britton, G., and Goodwin, T. W., Pergamon Press, New York, 1981.

Kootsey, J. M., Complexity and significance in computer simulations of physiological systems, FASEB, 46, 2490-2493, (1987).

Lehninger, A. L., Biochemistry, 2nd edition, Worth Publishers, New York, 1975.

Mavrovouniotis, M. L., Computer-aided design of biochemical pathways, Ph.D. Thesis, Department of Chemical Engineering, Massachusetts Institute of Technology, 1988.

Mavrovouniotis, M. L., Stephanopoulos, G. and Stephanopoulos, G., Formal modeling of approximate relations in biochemical systems, Biotechnol. Bioeng., 34, 196-206, (1989).

Miller, M. W., Yoneyama, M. and Soneda, M., *Phaffia*, a new yeast genus in the *Deuteromycotina* (*Blastomycetes*), Int. J. Sys. Bact, 26, 286-291, (1976).

Minkevich, I. G. and Eroshin, V. K., Productivity and heat generation of fermentation under oxygen limitation, Folia Microbiol., 18, 386-385, (1973).

Mistry, F. R., and Cooney, C. L., Production of ethanol by *Clostridium thermosaccharolyticum*: II. A quantitative model describing product distributions, Biotechnol. Bioeng., 34, 1305-1320, (1989).

Murtaugh, B. A. and Saunders, M. A., MINOS 5.1 User's Guide, Technical Report SOL 83-20R, Stanford University, 1987.

Noble, B. and Daniel, J. W., Applied Linear Algebra, Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1988.

Okagbue, R. N. and Lewis M. L., Influence of mixed culture conditions on yeast-wall hydrolytic activity of *Bacillus circulans* WL-12 and on extractability of astaxanthin from the yeast *Phaffia rhodozyma*, J. App. Bact., 59, 243-255, (1985).

Palsson, B. O., Joshi, A. and Ozturk, S. S., Reducing complexity in metabolic networks: making metabolic meshes manageable, FASEB, 46, 2485-2489, (1987).

Papoutsakis, E. T. and Meyer, C. L., Equations and calculations of product yields and preferred pathways for butanediol and mixed-acid fermentations, Biotechnol. Bioeng., 27, 50-66, (1985a).

Papoutsakis, E. T. and Meyer, C. L., Fermentation equations for propionic-acid bacteria and production of assorted oxychemicals from various sugars, Biotechnol. Bioeng., 27, 67-80, (1985b).

Papoutsakis, E. T., Equations and calculations for fermentations of butyric acid bacteria, Biotechnol. Bioeng., 26, 174-187, (1984).

Ramkrishna, D., A cybernetic perspective of microbial growth, 161-178, in Foundations of Biochemical Engineering: Kinetics and Thermodynamics in Biological Systems, edited by Blanch, H. W.,

Papoutsakis, E. T., and Stephanopoulos, G., American Chemical Society, 1983.

Reardon, K. F., Scheper, T. H., and Bailey, J. E., Metabolic pathway rates and culture fluorescence in batch fermentations of Clostridium acetobutylicum, Biotech. Prog., 3, 153-167, (1987).

Roels, J. A., Application of macroscopic principles to microbial metabolism, Biotechnol. Bioeng., 22, 2457-2514, (1980).

Roels, J. A., Energetics and Kinetics in Biotechnology, Elsevier Biomedical Press, New York, 1983.

Schneider, D. R., and Reklaitis, G. V., On material balances for chemically reacting systems, Chem. Eng. Sci., 30, 243-247, (1975).

Schrage, L. User's Manual for Linear, Integer, and Quadratic Programming with LINDO, 4th Ed., The Scientific Press, Redwood City, California, 1989.

Schubert, E., and Hofmann, H., Reaction engineering. 1. Stoichiometry of complex reactions, Int. Chem. Eng., 16, 132-136, (1976).

Sedgewick, R. Algorithms, Addison-Wesley Publishing Company, Inc., Reading, Massachusetts., 1984.

Seressiotis, A. and Bailey, J. E., MPS: An algorithm and data base for metabolic pathway synthesis, Biotechnol. Lett., 8, 837-842, (1986).

Seressiotis, A. and Bailey, J. E., MPS: An artificially intelligent software system for the analysis and synthesis of metabolic pathways, Biotechnol. Bioeng., 31, 587-602, (1988).

Seressiotis, A., Mathematical models based on molecular processes of the cell, Ph.D. Thesis, Department of Chemical Engineering, California Institute of Technology, 1988.

Shu, J. and Shuler, M. L., A mathematical model for the growth of a single cell of *E. coli* on a glucose/glutamine/ammonium medium, Biotechnol. Bioeng., 33, 1117-1126, (1989).

Simpson, K. L. and Chichester, C. O., Carotenoid pigments of yeasts, 493-515, in The Yeasts, edited by A. H. Rose and J. S. Harrison, Academic Press, New York, 1971.

Steinmeyer, D. E. and Shuler, M. L., Structured model for *Saccharomyces cerevisiae*, Chem. Eng. Sci., 44, 2017-2030, (1989).

Stephanopoulos, G. and Stephanopoulos, G., Artificial intelligence in the development and design of biochemical processes, Trends Biotechnol., 4, 241-249, (1986).

Stephanopoulos, G., and San, K., Studies on on-line bioreactor identification. I. Theory, Biotechnol. Bioeng., 26, 1176-1188, (1984).

Stouthamer, A. H., A theoretical study on the amount of ATP required for synthesis of microbial cell material, Antonie van Leeuwenhoek, 39, 545-565, (1973).

Stouthamer, A. H., Theoretical calculations on the influence of the inorganic nitrogen source on parameters for aerobic growth of microorganisms, Antonie van Leeuwenhoek, 43, 351-367, (1977).

Strang, G., Linear Algebra and Its Applications, Academic Press, New York, 1980.

Stryer, L. B., Biochemistry, 3d edition, W. H. Freeman and Company, New York, 1988.

Swartz, J. R. and Cooney, C. L., Indirect fermentation measurements as a basis for control, Biotechnol. Bioeng. Symp., 9, 95-101, (1979).

Sørensen, J. P. and Stewart, W. E., Structural analysis of multi-component reaction models, AIChE J., 26, 98-111, (1980).

Thauer, R. K., Jungermann, K., and Decker, K., Energy conservation in chemotrophic anaerobic bacteria, Bact. Rev., 41, 101-180, (1977).

Tsai, S. P. and Lee, Y. H., Application of Gibbs' rule and a simple pathway method to microbial stoichiometry, Biotech. Prog., 4, 82-88, (1988).

Turner, B. G., Ramkrishna, D., and Jansen, N. B., Cybernetic modeling of bacterial cultures at low growth rates: single-substrate systems, Biotechnol. Bioeng., 34, 252-261, (1989).

Verhoff, F. H., and Spradlin, J. E., Mass and energy balance analysis of metabolic pathways applied to citric acid production by *Aspergillus niger*, Biotechnol. Bioeng., 18, 425-432, (1976).

Wang, H. Y., Cooney, C. L., and Wang, D. I. C., Computer-aided baker's yeast fermentations, Biotechnol. Bioeng., 19, 69-86, (1977).

Wang, H. Y., Cooney, C. L., and Wang, D. I. C., On-line gas analysis for material balances and control, Biotechnol. Bioeng. Symp., 9, 13-23, (1979).

Wang, N. S., and Stephanopoulos, G., Application of macroscopic balances to the identification of gross measurement errors, Biotechnol. Bioeng., 25, 2177-2208, (1983).

Welch, G. R., Some problems in the usage of Gibbs free energy in biochemistry, J. Theor. Biol., 114, 433-446, (1985).

Whitwell, J. C., and Dartt, S. R., Independent reactions in the presence of isomers, AIChE J.,19, 1114-1120, (1973).